Semi-Blind Channel-and-Signal Estimation for Uplink Massive MIMO With Channel Sparsity

WENJING YAN AND XIAOJUN YUAN[ID]**, (Senior Member, IEEE)**
Center for Intelligent Networking and Communications, National Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China

Corresponding author: Xiaojun Yuan (xjyuan@uestc.edu.cn)

ABSTRACT This paper considers the transceiver design for uplink massive multiple-input multiple-output (MIMO) systems with channel sparsity in the angular domain. Recent progress has shown that sparsity-learning-based blind signal detection is able to retrieve the channel and data by using massage-passing-based sparse matrix factorization methods. Short pilot sequences are inserted into user packets to eliminate the so-called phase and permutation ambiguities inherent in sparse matrix factorization. In this paper, to exploit the knowledge of these short pilot sequences more efficiently, we propose a semi-blind channel-and-signal estimation (SCSE) scheme in which the knowledge of the pilot sequences are integrated into the message passing algorithm for sparse matrix factorization. The SCSE algorithm involves enumeration over all possible user permutations, and so is time-consuming when the number of users is relatively large. To reduce complexity, we further develop the simplified SCSE (S-SCSE) to accommodate systems with a large number of users. The numerical results show that our semi-blind signal detection scheme substantially outperforms the state-of-the-art blind detection and training-based schemes in the short-pilot regime.

INDEX TERMS Massive MIMO, channel sparsity, semi-blind, message passing, training-based, blind detection.

I. INTRODUCTION

Massive multiple-input multiple-output (MIMO) systems [1]–[4] achieve significant performance improvement over the traditional communication systems in many aspects, such as increasing channel capacity, suppressing channel fading, and enhancing energy efficiency, etc. In a massive MIMO scenario, a base station (BS) typically equipped with an array of a few hundred antennas simultaneously serves many tens of terminals in a single time-frequency resource slot [5]. As the scale of the terminals or the array increases, the acquisition of channel state information (CSI) becomes one of the key obstacles for the utilization of massive MIMO [6].

Many studies have been attracted to the design of efficient and reliable techniques for channel acquisition. We are particularly focused on the uplink case, where users transmit signals to a BS [7]. In a training-based approach, each transmission frame is divided into two phases, namely, a training phase and a data transmission phase [7], [8]. In the train-

ing phase, pilots are transmitted to facilitate the estimation of the channel coefficients at the receiver side; in the data transmission phase, data are transmitted and the receiver performs detection based on the estimated channel.

Compared to separate signal processing for the two phases, joint channel-and-signal estimation is able to improve the system performance since partially detected data can be used as *soft* pilots to enhance the channel estimation accuracy in an iterative fashion [9]. However, no matter whether separate or joint signal processing is employed, it is required in the training-based approach that the number of pilot symbols is no less than that of users, so as to ensure a vanishing channel estimation error [8], [10]. As such, channel acquisition generally consumes a substantial portion of the system resource.

To reduce the channel acquisition overhead, another line of research is called the blind detection approach, in which the channel and data were estimated with little prior information of the signals from the transmitter side [11]–[13]. In particular, it has been recently evidenced that a massive MIMO system exhibits channel sparsity in the angular domain, since signals usually impinge upon a massive antenna array from a limited range of angles [14]–[17]. Based on the channel

sparsity, a blind iterative detection technique [18] has been developed to avoid the use of pilots in channel acquisition. Approximate message passing algorithms [19], [20] are used to simultaneously estimate the channel and data by factorizing the received noisy observation matrix.

Sparsity-learning based blind detection in [18], however, can be improved in a number of aspects. For example, the blind detection scheme in [18] imposes a relatively stringent requirement on the channel sparsity level and the signal-to-noise ratio (SNR) of the system to achieve a satisfactory error performance. More importantly, blind detection suffers from the so-called phase and permutation ambiguities inherent in matrix factorization. In [18], a reference symbol and a user label are inserted in each user packet to eliminate the phase and permutation ambiguities after matrix factorization. Yet, as the reference symbols and the user labels (similar to pilots) are *a priori* known by the receiver, such knowledge can be integrated into the iterative process of matrix factorization to enhance the detection performance, rather than used for afterwards compensation.

To address the above issues, we propose a semi-blind detection scheme to jointly estimate the channel and the user signals in a sparse massive MIMO system. We focus on the scenario that a short pilot sequence is inserted into each user packet to assist the matrix factorization for joint channel and signal estimation. Here "short pilot" means that the pilot sequence is not long enough to generate a relatively accurate initial channel estimate, and so the existing training-based approaches [7], [9] are unable to achieve a good performance. We show that, to efficiently exploit the short pilots, the phase and permutation ambiguities need to be skilfully estimated in the iterative process of the matrix factorization. As such, a message-passing based semi-blind channel and signal estimation (SCSE) algorithm is developed, building upon the framework of approximate message passing algorithms. The main contributions of this paper are summarized as follows.

- *SCSE algorithm for massive MIMO:* We propose a novel semi-blind detection scheme for massive MIMO systems to jointly estimate the channel and the signals. The proposed semi-blind detection scheme is able to efficiently exploit the information of short pilots in the iterative process of sparse matrix factorization.
- *Simplified SCSE algorithm for complexity reduction:* The SCSE algorithm involves an exhaustive search of all possible user permutations, which is computationally infeasible when the number of users is relatively large. Thus, we develop a simplified SCSE (S-SCSE) algorithm to avoid the burden of permutation enumeration.
- We show that our proposed S-SCSE is able to substantially outperform the state-of-the-art training-based and blind detection approaches [9], [18] in the short-pilot regime. We also show that, compared to the SCSE algorithm, the S-SCSE algorithm significantly reduces the computational complexity while maintaining a

similar performance, thereby striking a favorable balance between complexity and performance.

The remainder of this paper is organized as follows: Section II describes the sparse channel model and the system model for uplink MIMO systems. In Section III, we formulate the joint channel and signal inference problem by including the estimation of the phase and permutation ambiguities inherent in sparse matrix factorization. Based on the formulated problem in Section III, the SCSE and S-SCSE algorithms are derived based on the message-passing principles in Section IV. Besides, the parameter tuning method based on the expectation maximization (EM) algorithm [21], the computational complexity comparisons between our proposed algorithms and the existing approaches, and the selection metric for random initializations are also described in Section IV. Numerical results are presented in Section V to verify the effectiveness of our proposed algorithms. Finally, we conclude the paper in Section VI.

Notations: Capital bold letters, lowercase bold letters, and regular letters represents matrices, vectors, and scalars, respectively. For any matrix $\mathbf{A}$, $\mathbf{a}_i$ refers to the ith column of $\mathbf{A}$, and $a_{i,j}$ refers to the (i,j)th entry of $\mathbf{A}$. $\mathbb{C}$ denotes the complex field; $\mathbb{R}$ denotes the real field; $\mathcal{S}$ denotes a set; $\mathbf{P}$ denotes an arbitrary permutation matrix. For any set $\mathcal{S}$, $|\mathcal{S}|$ represents the cardinality of $\mathcal{S}$; $\mathbf{e}_i = [0, \ldots, 0, 1, 0, \ldots, 0]^{\mathrm{T}}$ with the only non-zero element being at the ith position; for any scalar x, $|x|$ represents the absolute value of x; $\| \cdot \|_2$ represents the ℓ_2-norm; $\| \cdot \|_{\mathrm{F}}$ represents the Frobenius norm. The superscripts $(\cdot)^{\mathrm{T}}$, $(\cdot)^{\mathrm{H}}$, $(\cdot)^{-1}$ represent the transpose, the conjugate transpose, and the inverse of a matrix, respectively; $\mathbb{E}(\cdot)$, $\delta(\cdot)$ and $e^{(\cdot)}$ represent the expectation, the Dirichlet function, and the exponential function; diag$\{\mathbf{a}\}$ represents the diagonal matrix with the diagonal specified by $\mathbf{a}$; $\lceil a \rceil$ represents the minimum integer larger than a. For any integer $\mathcal{I}_N$ denotes the set of integers from 1 to N. $\mathcal{CN}(\cdot, \mu, \nu)$ represents a complex circularly symmetric Gaussian distribution with the mean μ and covariance ν.

II. SYSTEM MODEL
A. SPARSE CHANNEL MODEL

Consider an uplink massive MIMO system with K single-antenna transmitters and a receiver equipped with N antennas deployed as a uniform linear array (ULA), where $N \gg K \gg 1$. Denote by $\theta_{\ell,k}$ the angle of arrival (AoA) of the ℓth path from transmitter k. The array steering vector for an incident signal from angle $\theta_{\ell,k}$ can be written as

$$\mathbf{a}_r(\theta_{\ell,k}) = \left[1, e^{-j2\pi \frac{d}{\lambda} \cos(\theta_{\ell,k})}, \ldots, e^{-j2\pi \frac{(N-1)d}{\lambda} \cos(\theta_{\ell,k})} \right]^{\mathrm{T}} \quad (1)$$

where d is the interval between any two adjacent receive antennas, and λ is the wavelength of propagation. We use the virtual representation method in [14] to divide the signal AoAs into N resolution bins with the kth bin represented by $\theta_{\ell,k} = \arccos(\frac{\ell\lambda}{dN})$, $\ell \in \mathcal{I}_N \triangleq \{0, 1, \ldots, N-1\}$. Then, the physical channel of a massive MIMO system can be modeled as

$$\check{\mathbf{H}} = \mathbf{A}_r \mathbf{H} \quad (2)$$

where $\mathbf{A}_r = \frac{1}{\sqrt{N}}\left[\mathbf{1}, \mathbf{a}_r(\arccos\frac{\lambda}{dN}), \ldots, \mathbf{a}_r(\arccos\frac{(N-1)\lambda}{dN})\right] \in \mathbb{C}^{N\times N}$ is the discrete Fourier transform (DFT) matrix, $\mathbf{1}$ is a K-dimensional all-one vector, and $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_K]$ is the projection of the physical channel in the angular domain. Note that each $\mathbf{h}_k = [h_{1,k}, h_{2,k}, \ldots, h_{N,k}]^T$ is the complex channel coefficient vector of user k, where $h_{n,k}$ is the aggregate channel coefficient in the resolution bin centered around $\theta_{\ell,k}$.

The physical channel of a massive MIMO system exhibits a sparse structure in the angular domain, since only a small portion of resolution bins receive electromagnetic waves from the transmitters. Therefore, the angular-domain channel representation $\mathbf{H}$ is a sparse matrix with a large portion of the elements being zero or close to zero [14]. Define the sparsity level of the massive MIMO channel by

$$\rho = \frac{|\mathcal{S}|}{NK} < 1 \tag{3}$$

where $\mathcal{S}$ is the support of $\mathbf{H}$.

We assume that each entry of $\mathbf{h}_k$ is independently drawn from a Bernoulli circularly symmetric complex Gaussian (B-CSCG) distribution $(1-\rho)\delta(h) + \rho\mathcal{CN}(h; 0, \sigma_{h,k}^2)$, where $\delta(\cdot)$ is the Dirac delta function, and $\sigma_{h,k}^2$ is the average power of the non-zero coefficients of the channel $\mathbf{h}_k$. Note that $\sigma_{h,k}^2$ is determined by the large-scale fading of user k, and is generally unknown to the receiver.[1] Based on the above discussion, the distribution of the channel is given by

$$P_\mathbf{H}(\mathbf{H}) = \prod_{n=1}^N \prod_{k=1}^K (1-\rho)\delta(h_{n,k}) + \rho\mathcal{CN}(h_{n,k}; 0, \sigma_{h,k}^2). \tag{4}$$

B. SYSTEM MODEL
Assume that the channel is block-fading with coherence time T. The massive MIMO system in the angular domain over T time slots can be modeled as

$$\mathbf{Y} = \mathbf{HX} + \mathbf{W} = \mathbf{Z} + \mathbf{W} \tag{5}$$

where $\mathbf{Y} \in \mathbb{C}^{N\times T}$ is the transformed observation matrix in the angular domain, $\mathbf{W} \in \mathbb{C}^{N\times T}$ is an additive white Gaussian noise (AWGN) with each entry independently drawn from $\mathcal{CN}(0, N_0)$ with N_0 being the noisy power, $\mathbf{H} \in \mathbb{C}^{N\times K}$ is the angular-domain channel matrix as aforementioned, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_K]^T \in \mathbb{C}^{K\times T}$ is the signal matrix, and $\mathbf{Z} = \mathbf{HX} \in \mathbb{C}^{N\times T}$. Each entry of $\mathbf{X}$ is modulated by using a constellation $\mathcal{C} = \{c_1, c_2, \ldots, c_{|\mathcal{C}|}\}$, where $|\mathcal{C}|$ is the cardinality of $\mathcal{C}$. That is, $x_{k,t}$ is uniformly drawn from $\mathcal{C}$ for $\forall k, t$, where $x_{k,t}$ is the tth entry of $\mathbf{x}_k$. Assume that $\mathcal{C}$ is rotationally invariant for any rotation angle $\theta \in \Omega$, where $\Omega = \{\omega_1, \omega_2, \ldots, \omega_{|\Omega|}\}$ is an angle set. For example, $\Omega = \{0°, 90°, 180°, 270°\}$ for standard quadrature amplitude modulation (QAM). For each user k, the first T_P symbols of $\mathbf{x}_k$ (denoted by $\mathbf{x}_{P,k} \in \mathbb{C}^{P\times 1}$) are assigned as pilots,

and the remaining $T - T_P$ are data symbols. We use $\mathcal{T}_P$ to represent the set $\{1, 2, \ldots, T_P\}$, and $\mathcal{T}_D$ to represent the set $\{T_P+1, T_P+2, \ldots, T\}$. Let $\mathbf{X}_P = [\mathbf{x}_{P,1}, \mathbf{x}_{P,2}, \ldots, \mathbf{x}_{P,K}]^T$ be the pilot matrix occupying the first T_P columns of $\mathbf{X}$, and let $\mathbf{X}_D = [\mathbf{x}_{D,1}, \mathbf{x}_{D,2}, \ldots, \mathbf{x}_{D,K}]^T$ be the data matrix occupying the remaining $T - T_P$ columns of $\mathbf{X}$, i.e., $\mathbf{X} = [\mathbf{X}_P, \mathbf{X}_D]$. Similarly, $\mathbf{Y}$ can be expressed as $\mathbf{Y} = [\mathbf{Y}_P, \mathbf{Y}_D]$, where $\mathbf{Y}_P$ and $\mathbf{Y}_D$ correspond to $\mathbf{X}_P$ and $\mathbf{X}_D$, respectively. Assume that the entries of $\mathbf{X}$ are independent of each other, i.e.,

$$P_\mathbf{X}(\mathbf{X}) = \prod_{k=1}^K \prod_{t=1}^T p_{x_{k,t}}(x_{k,t}). \tag{6}$$

Denote by P the total power budget of the transmitters and $\alpha_k P$ the average transmission power of the kth transmitter. Then, each transmitter is power-constrained as

$$\frac{1}{T}\mathbb{E}[\mathbf{x}_k^H \mathbf{x}_k] \leqslant P_k, \quad \text{for all } k \in \mathcal{I}_K \triangleq \{1, 2, \ldots, K\} \tag{7}$$

where P_k is the power upper bound of user k.

III. PROBLEM DESCRIPTION
In this paper, our goal is to retrieve both the channel matrix $\mathbf{H}$ and the symbol matrix $\mathbf{X}_D$ from the observed data matrix $\mathbf{Y}$ together with the prior knowledge of the pilot matrix $\mathbf{X}_P$. This problem can be formulated by using the *maximum a posteriori* (MAP) principle as

$$\left(\hat{\mathbf{H}}, \hat{\mathbf{X}}_D\right) = \arg\max_{\mathbf{H}, \mathbf{X}_D} p_{\mathbf{H}, \mathbf{X}_D|\mathbf{Y}, \mathbf{X}_P}(\mathbf{H}, \mathbf{X}_D|\mathbf{Y}, \mathbf{X}_P) \tag{8}$$

where $\hat{\mathbf{H}}$ and $\hat{\mathbf{X}}_D$ are the estimates of the channel matrix $\mathbf{H}$ and the signal matrix $\mathbf{X}_D$, respectively.

A straightforward approach to solving (8) is first to estimate the channel $\mathbf{H}$ based on $\mathbf{Y}_P$ (with known $\mathbf{X}_P$) and then to estimate the data matrix $\mathbf{X}_D$ based on $\mathbf{Y}_D$ and the channel estimate $\hat{\mathbf{H}}$. This approach is referred to as the separate channel-and-signal estimation method scheme in the following. In principle, the estimated data can be used to further refine the channel estimate and hence improve the system performance. To this end, the authors in [9] proposed a joint channel and signal estimation method which involves approximate message passing over the factor graph obtained by factorizing the probability distribution $p_{\mathbf{H}, \mathbf{X}_D|\mathbf{Y}, \mathbf{X}_P}$ in (8).

In this paper, we focus on the ''semi-blind'' scenario where the pilot length T_P is not large enough to provide a relatively accurate initial channel estimate for data detection. In this scenario, both the separate and joint training-based estimation approaches discussed above do not work well. The main reason is that when T_P is small, the scheme in [9] is close to blind channel-and-signal estimation in which the issue of phase and permutation ambiguities arise [18]. For a small T_P, the knowledge of $\mathbf{X}_P$ is not ''strong'' enough to correct the phase and permutation ambiguities in the iterative estimation process. As a result, the scheme in [9] may perform even worse than the blind detection method in [18].

Instead, we aim to find a method that can efficiently exploit the knowledge of $\mathbf{X}_P$ in the presence of phase and

---

[1]We assume that $\{\sigma_{h,k}^2\}$ are independently drawn from a certain known distribution. The distribution will be specified in the simulations in Section V. The estimation of $\{\sigma_{h,k}^2\}$ will be discussed later in Section IV-D.

permutation ambiguities. To this end, we need to estimate the phase and permutation ambiguities in the iterative process of message passing. Denote by $\Pi = [\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots, \boldsymbol{\pi}_K]^T \in \mathbb{R}^{K \times K}$ an arbitrary permutation matrix, and by $\Sigma = \text{diag}\{\sigma_1, \sigma_2, \ldots, \sigma_K\}$ a diagonal matrix where each phase shift $\sigma_i = e^{j\omega_i}$ with ω_i independently and uniformly taken from Ω. It is known in [18] that blind detection suffers from the phase and permutation ambiguities, i.e., if $(\hat{\mathbf{H}}, \hat{\mathbf{X}})$ is a valid factorization given $\mathbf{Y}$ in (5), then $(\hat{\mathbf{H}}\Pi^{-1}\Sigma^{-1}, \Sigma\Pi\hat{\mathbf{X}})$ is also a valid factorization. We will estimate Π and Σ in the message passing process by exploiting the knowledge of $\mathbf{X}_P$. To this end, we define auxiliary variables

$$\tilde{\mathbf{H}} = \mathbf{H}\Pi^{-1}\Sigma^{-1} \quad \text{and} \quad \tilde{\mathbf{X}} = \Sigma\Pi\mathbf{X}. \tag{9}$$

Recall from (4) that the columns of $\mathbf{H}$ are independent and identically distributed (i.i.d.). Thus, for any permutation Π, $\mathbf{H}\Pi^{-1}$ has the same distribution as $\mathbf{H}$ does. Then,

$$p(\mathbf{H}\Pi^{-1}, \Pi) = p(\mathbf{H}\Pi^{-1}|\Pi)p(\Pi) = p(\mathbf{H}\Pi^{-1})p(\Pi). \tag{10}$$

That is, $\mathbf{H}\Pi^{-1}$ is independently of Π. Similarly, for any phase ambiguity matrix Σ, $\tilde{\mathbf{H}} = \mathbf{H}\Pi^{-1}\Sigma^{-1}$ has the same distribution as $\mathbf{H}$ does. Therefore, $\tilde{\mathbf{X}}$ is independent of Π and Σ. Since $\mathbf{H}$, $\mathbf{X}$, Π, and Σ are independent of each other, we conclude that $\tilde{\mathbf{H}}$, $\tilde{\mathbf{X}}$, Π, and Σ are also independent of each other, i.e.,

$$p_{\tilde{\mathbf{H}}, \tilde{\mathbf{X}}, \Pi, \Sigma}(\tilde{\mathbf{H}}, \tilde{\mathbf{X}}, \Pi, \Sigma) = p_{\tilde{\mathbf{H}}}(\tilde{\mathbf{H}})p_{\tilde{\mathbf{X}}}(\tilde{\mathbf{X}})p_\Pi(\Pi)p_\Sigma(\Sigma). \tag{11}$$

Then we recast the problem in (8) as

$$\max_{\tilde{\mathbf{H}}, \tilde{\mathbf{X}}, \Sigma, \Pi} p_{\tilde{\mathbf{H}}, \tilde{\mathbf{X}}, \Sigma, \Pi|\mathbf{Y}, \mathbf{X}_P}(\tilde{\mathbf{H}}, \tilde{\mathbf{X}}, \Sigma, \Pi|\mathbf{Y}, \mathbf{X}_P). \tag{12}$$

From the Bayes' rule, we obtain

$$p_{\tilde{\mathbf{H}}, \tilde{\mathbf{X}}, \Sigma, \Pi|\mathbf{Y}, \mathbf{X}_P}(\tilde{\mathbf{H}}, \tilde{\mathbf{X}}, \Sigma, \Pi|\mathbf{Y}, \mathbf{X}_P)$$

$$= \frac{p_{\mathbf{Y}, \mathbf{X}_P, \tilde{\mathbf{H}}, \tilde{\mathbf{X}}, \Sigma, \Pi}(\mathbf{Y}, \mathbf{X}_P, \tilde{\mathbf{H}}, \tilde{\mathbf{X}}, \Sigma, \Pi)}{p_{\mathbf{Y}, \mathbf{X}_P}(\mathbf{Y}, \mathbf{X}_P)} \tag{13a}$$

$$\propto p_{\mathbf{Y}|\mathbf{X}_P, \tilde{\mathbf{H}}, \tilde{\mathbf{X}}, \Sigma, \Pi}(\mathbf{Y}|\mathbf{X}_P, \tilde{\mathbf{H}}, \tilde{\mathbf{X}}, \Sigma, \Pi)$$
$$\times p_{\mathbf{X}_P|\tilde{\mathbf{H}}, \tilde{\mathbf{X}}, \Sigma, \Pi}(\mathbf{X}_P|\tilde{\mathbf{H}}, \tilde{\mathbf{X}}, \Sigma, \Pi)p_{\tilde{\mathbf{H}}, \tilde{\mathbf{X}}, \Sigma, \Pi}(\tilde{\mathbf{H}}, \tilde{\mathbf{X}}, \Sigma, \Pi) \tag{13b}$$

$$= p_{\mathbf{Y}|\tilde{\mathbf{H}}, \tilde{\mathbf{X}}}(\mathbf{Y}|\tilde{\mathbf{H}}, \tilde{\mathbf{X}})p_{\mathbf{X}_P|\tilde{\mathbf{X}}, \Sigma, \Pi}(\mathbf{X}_P|\tilde{\mathbf{X}}, \Sigma, \Pi)$$
$$\times p_{\tilde{\mathbf{H}}}(\tilde{\mathbf{H}})p_{\tilde{\mathbf{X}}}(\tilde{\mathbf{X}})p_\Sigma(\Sigma)p_\Pi(\Pi) \tag{13c}$$

$$= \left[\prod_n \prod_t p_{y_{n,t}|z_{n,t}} \left(y_{n,t}|z_{n,t} = \sum_k \tilde{h}_{n,k}\tilde{x}_{k,t} \right) \right]$$
$$\times \left[\prod_n \prod_k p_{\tilde{h}_{n,k}}(\tilde{h}_{n,k}) \right] \left[\prod_k \prod_t p_{\tilde{x}_{k,t}}(\tilde{x}_{k,t}) \right]$$
$$\times \delta(\tilde{\mathbf{X}}_P - \Sigma\Pi\mathbf{X}_P) \left[\prod_k p_{\sigma_k}(\sigma_k) \right] p_\Pi(\Pi) \tag{13d}$$

where (13a) follows from the Bayes' rule; the notation $\propto$ in step (13b) means equality up to a constant scaling factor; (13c) is from the facts that (i) $(\tilde{\mathbf{X}}_P, \Sigma, \Pi) \to (\tilde{\mathbf{H}}, \tilde{\mathbf{X}}) \to \mathbf{Y}$ forms a Markov chain, (ii) $\mathbf{X}_P$ is independent of $\tilde{\mathbf{H}}$ for any given $\tilde{\mathbf{X}}$, Σ, and Π, and (iii) $\tilde{\mathbf{H}}$, $\tilde{\mathbf{X}}$, Σ, Π are independent

of each other by (11); $p_{\sigma_k}(\sigma_k)$ in (13d) is the probability density of the phase shift of user k. Recall that $p_{\sigma_k}(\sigma_k) = \frac{1}{|\Omega|} \sum_{\omega \in \Omega} \delta(\sigma_k - e^{j\omega})$, where $|\Omega|$ is the cardinality of Ω. The factorization in (13) will be used in the development of message passing algorithms in the subsequent section. We will show that the message passing algorithms developed based on (13) performs much better than the one based on the factorization of $p_{\mathbf{H}, \mathbf{X}_D|\mathbf{Y}, \mathbf{X}_P}$ as in [9]

IV. SEMI-BLIND MESSAGE PASSING ALGORITHMS
A. FACTOR GRAPH REPRESENTATION
To describe the message passing process more clearly, we introduce an auxiliary variable $\mathbf{P} \in \mathbb{R}^{K \times K}$ to denote a random permutation matrix, where $\mathbf{P} \in \mathcal{P} \triangleq \{\mathbf{P}_1, \mathbf{P}_2, \cdots, \mathbf{P}_{K!}\}$ with an equal probability with $\mathcal{P}$ being the set of all permutations.[2] Then, the fact that Π is a random permutation can be represented by the following joint distribution:

$$p_{\Pi, \mathbf{P}}(\Pi, \mathbf{P}) = p_{\mathbf{P}}(\mathbf{P})\delta\left([\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots, \boldsymbol{\pi}_K]^T - \mathbf{P} \right) \tag{14}$$

where $p_{\mathbf{P}}(\mathbf{P}) = \frac{1}{k!} \sum_{\ell=1}^{K!} \delta(\mathbf{P} - \mathbf{P}_\ell)$, $\boldsymbol{\pi}_k^T$ is taken from the set $\{\mathbf{e}_i\}_{i=1}^K$ with $\mathbf{e}_i = [0, \ldots, 0, 1, 0, \ldots, 0]^T$ being the ith column of the K-by-K identity matrix. With the inclusion of the auxiliary variable $\mathbf{P}$, the factorization in (13) converts to

$$p_{\tilde{\mathbf{H}}, \tilde{\mathbf{X}}, \Sigma, \Pi, \mathbf{P}|\mathbf{Y}, \mathbf{X}_P}(\tilde{\mathbf{H}}, \tilde{\mathbf{X}}, \Sigma, \Pi, \mathbf{P}|\mathbf{Y}, \mathbf{X}_P)$$

$$= \left[\prod_n \prod_t p_{y_{n,t}|z_{n,t}} \left(y_{n,t}|z_{n,t} = \sum_k \tilde{h}_{n,k}\tilde{x}_{k,t} \right) \right]$$
$$\times \left[\prod_n \prod_k p_{\tilde{h}_{n,k}}(\tilde{h}_{n,k}) \right] \left[\prod_k \prod_t p_{\tilde{x}_{k,t}}(\tilde{x}_{k,t}) \right]$$
$$\times \left[\prod_{k=1}^K \delta(\tilde{\mathbf{x}}_{P,k} - \sigma_k\boldsymbol{\pi}_k^T\mathbf{X}_P) \right] \left[\prod_k p_{\sigma_k}(\sigma_k) \right]$$
$$\times p_{\mathbf{P}}(\mathbf{P})\delta\left([\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots, \boldsymbol{\pi}_K]^T - \mathbf{P} \right). \tag{15}$$

The factorized posterior distribution in (15) can be represented by a factor graph, as depicted in Fig. 1. In Fig. 1, we use a brief form of δ_k to represent $\delta(\tilde{\mathbf{x}}_{P,k} - \sigma_k\boldsymbol{\pi}_k^T\mathbf{X}_P)$, $k \in \mathcal{I}_K$, and δ_Π to represent $\delta([\boldsymbol{\pi}_1, \boldsymbol{\pi}, \ldots, \boldsymbol{\pi}_K]^T - \mathbf{P})$. Each hollow circle in Fig. 1 represents a "variable node" corresponding to a random variable involved in (15), and each black solid square represents a "factor node" corresponding to a factor function in (15). A variable node is connected to a factor node if the variable appears in the factor function.

In Fig. 1, we divide the whole factor graph into two parts. In part I, we estimate the channel matrix $\tilde{\mathbf{H}}$ and signal matrix $\tilde{\mathbf{X}}$ based on $\mathbf{Y}$ and the knowledge that $\tilde{\mathbf{H}}$ is sparse; in part II, we use the knowledge of $\mathbf{X}_P$ to improve the estimation of $\tilde{\mathbf{X}}_P$ and estimate Π and Σ based on the constraint of $\tilde{\mathbf{X}}_P = \Sigma\Pi\mathbf{X}_P$. We derive the semi-blind detection algorithm

---

[2]It is possible to design message passing directly based on the factorization in (12). However, the introduction of $\mathbf{P}$ yields a unified view of the derivations of the SCSE and S-SCSE algorithms. In particular, we show in Section IV-C that the S-SCSE algorithm is derived simply by deleting the constraint δ_Π in Fig. 1.
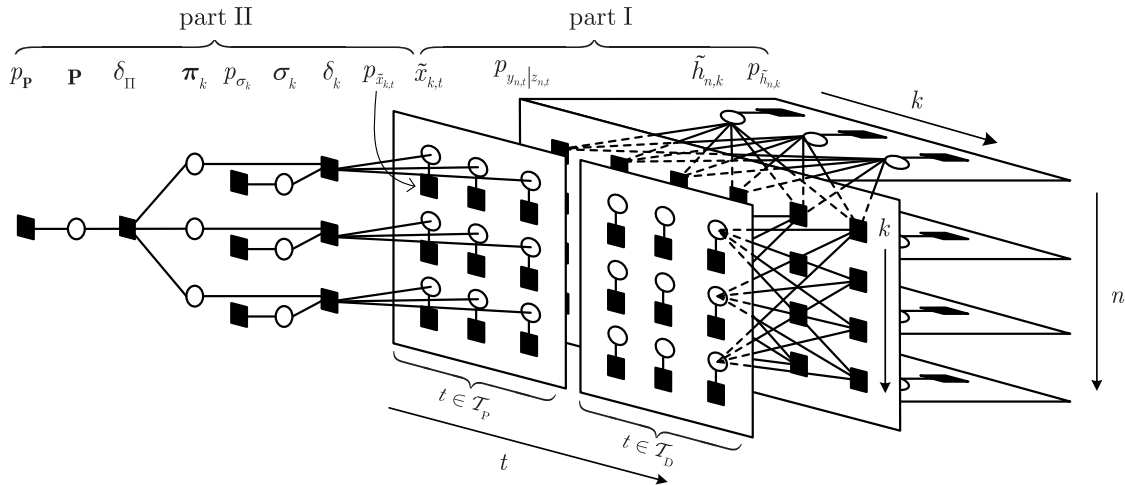
FIGURE 1. The factor graph representation for the joint probability in (15) with $N = 4$, $K = 3$, $T_P = 3$, and $T = 6$.

based on the message passing principles over the factor graph in Fig. 1. Note that the constraints in part I are related to factorizing the matrix product $\tilde{\mathbf{H}}\tilde{\mathbf{X}}$. This part can be realized by following the BiG-AMP algorithm in [20]. Therefore, in what follows, we focus on the derivation of the message passing algorithm for part II.

B. SCSE ALGORITHM

The messages involved in part II are described as follows. Denote by $\Delta_{a \to b}(\cdot)$ the message from node a to node b and by $\Delta_c(\cdot)$ the marginal posterior of variable c.

1) The message from $\tilde{x}_{k,t}$ to δ_k is given by

$$\Delta_{\tilde{x}_{k,t} \to \delta_k}(\tilde{x}_{k,t}) \propto p_{\tilde{x}_{k,t}}(\tilde{x}_{k,t}) \prod_{n=1}^{N} \Delta_{p_{y_{n,t}|z_{n,t}} \to \tilde{x}_{k,t}}(\tilde{x}_{k,t}). \quad (16)$$

2) The message from σ_k to δ_k is given by

$$\Delta_{\sigma_k \to \delta_k}(\sigma_k) = p_{\sigma_k}(\sigma_k) = \frac{1}{|\Omega|} \sum_{\omega \in \Omega} \delta(\sigma_k - e^{j\omega}). \quad (17)$$

3) The message from δ_k to $\boldsymbol{\pi}_k$ is given by

$$\Delta_{\delta_k \to \boldsymbol{\pi}_k}(\boldsymbol{\pi}_k)$$
$$\propto \int_{\sigma_k, \{x_{k,t}\}_{t=1}^{T_P}} \delta(\tilde{\mathbf{x}}_{\mathbf{P},k} - \sigma_k \boldsymbol{\pi}_k^{\mathrm{T}} \mathbf{X}_{\mathbf{P}})$$
$$\times p_{\sigma_k}(\sigma_k) \prod_{t=1}^{T_P} \Delta_{\tilde{x}_{k,t} \to \delta_k}(\tilde{x}_{k,t}). \quad (18)$$

Since $\tilde{x}_{k,t}$, $\boldsymbol{\pi}_k$, and σ_k are discrete variables, we write the message in (18) in its discrete form as

$$P_{\delta_k \to \boldsymbol{\pi}_k}(\boldsymbol{\pi}_k = \mathbf{e}_i)$$
$$= \frac{1}{C} \sum_{\omega \in \Omega} \prod_{t=1}^{T_P} P_{\tilde{x}_{k,t} \to \delta_k}(\tilde{x}_{k,t} = e^{j\omega} x_{i,t}), \quad i \in \mathcal{I}_K \quad (19)$$

where $P_{\tilde{x}_{k,t} \to \delta_k}(\tilde{x}_{k,t} = e^{j\omega} x_{i,t})$ denotes the probability of $\tilde{x}_{k,t} = e^{j\omega} x_{i,t}$ specified in the message $\Delta_{\tilde{x}_{k,t} \to \delta_k}(\tilde{x}_{k,t})$, and C is a generic normalization factor. Clearly, the message from $\boldsymbol{\pi}_k$ to δ_{Π} is

$$\Delta_{\boldsymbol{\pi}_k \to \delta_{\Pi}}(\boldsymbol{\pi}_k) = \Delta_{\delta_k \to \boldsymbol{\pi}_k}(\boldsymbol{\pi}_k). \quad (20)$$

4) The message from $\mathbf{P}$ to δ_{Π} is given by

$$\Delta_{\mathbf{P} \to \delta_{\Pi}}(\mathbf{P}) = p_{\mathbf{P}}(\mathbf{P}) = \frac{1}{K!} \sum_{\ell=1}^{K!} \delta(\mathbf{P} - \mathbf{P}_\ell). \quad (21)$$

5) Combining the message from $\mathbf{P}$ to δ_{Π} and the messages from $\{\boldsymbol{\pi}_{k'}\}_{k' \neq k}^K$ to δ_{Π}, we obtain

$$\Delta_{\delta_{\Pi} \to \boldsymbol{\pi}_k}(\boldsymbol{\pi}_k)$$
$$\propto \int_{\mathbf{P}, \{\boldsymbol{\pi}_{k'}\}_{k'=1 \neq k}^K} \delta([\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \dots, \boldsymbol{\pi}_K]^{\mathrm{T}} - \mathbf{P})$$
$$\times \prod_{k'=1 \neq k}^K \Delta_{\boldsymbol{\pi}_{k'} \to \delta_{\Pi}}(\boldsymbol{\pi}_{k'}) p_{\mathbf{P}}(\mathbf{P}). \quad (22)$$

Denote by $\mathbf{p}_{\ell,k}$ the transpose of the kth row of $\mathbf{P}_\ell$. The discrete form of the above message can be written as

$$P_{\delta_{\Pi} \to \boldsymbol{\pi}_k}(\boldsymbol{\pi}_k = \mathbf{e}_i)$$
$$= \frac{1}{C} \sum_{\substack{\ell=1 \\ \ell: \, \mathbf{p}_{\ell,k} = \mathbf{e}_i}}^{K!} \prod_{k'=1 \neq k}^K P_{\boldsymbol{\pi}_{k'} \to \delta_{\Pi}}(\boldsymbol{\pi}_{k'} = \mathbf{p}_{\ell,k'}), \quad i \in \mathcal{I}_K \quad (23)$$

where $P_{\boldsymbol{\pi}_{k'} \to \delta_{\Pi}}(\boldsymbol{\pi}_{k'} = \mathbf{p}_{\ell,k'})$ denotes the probability of $\boldsymbol{\pi}_{k'} = \mathbf{p}_{\ell,k'}$ specified by the message $\Delta_{\boldsymbol{\pi}_{k'} \to \delta_{\Pi}}(\boldsymbol{\pi}_{k'})$. Similar to (20), the message from $\boldsymbol{\pi}_k$ to δ_k is

$$\Delta_{\boldsymbol{\pi}_k \to \delta_k}(\boldsymbol{\pi}_k) = \Delta_{\delta_{\Pi} \to \boldsymbol{\pi}_k}(\boldsymbol{\pi}_k). \quad (24)$$

6) The message from δ_k to $\tilde{x}_{k,t}$ is given by

$$
\Delta_{\delta_k \to \tilde{x}_{k,t}}(\tilde{x}_{k,t})
$$
$$
\propto \int_{\boldsymbol{\pi}_k, \sigma_k, \{x_{k,t'}\}_{t'=1 \neq t}^{T_P}} \delta(\tilde{\mathbf{x}}_{P,k} - \sigma_k \boldsymbol{\pi}_k^T \mathbf{X}_P)
$$
$$
\times p_{\sigma_k}(\sigma_k) \Delta_{\boldsymbol{\pi}_k \to \delta_k}(\boldsymbol{\pi}_k) \prod_{t'=1 \neq t}^{T_P} \Delta_{\tilde{x}_{k,t'} \to \delta_k}(\tilde{x}_{k,t'}). \quad (25)
$$

We write the above message in its discrete form as

$$
P_{\delta_k \to \tilde{x}_{k,t}}(\tilde{x}_{k,t} = c)
$$
$$
= \frac{1}{C} \sum_{\substack{i \in \mathcal{I}_K, \omega \in \Omega \\ \text{for } e^{j\omega} x_{i,t} = c}} P_{\delta_\Pi \to \boldsymbol{\pi}_k}(\boldsymbol{\pi}_k = \mathbf{e}_i)
$$
$$
\times \prod_{t'=1 \neq t}^{T_P} P_{\tilde{x}_{k,t'} \to \delta_k}(\tilde{x}_{k,t'} = e^{j\omega} x_{i,t'}), \quad c \in \mathcal{C}. \quad (26)
$$

7) The message from $\tilde{x}_{k,t}$ to $p_{y_{n,t}|z_{n,t}}$ is given by

$$
\Delta_{\tilde{x}_{k,t} \to p_{y_{n,t}|z_{n,t}}}(\tilde{x}_{k,t})
$$
$$
\propto p_{\tilde{x}_{k,t}}(\tilde{x}_{k,t}) \Delta_{\delta_k \to \tilde{x}_{k,t}}(\tilde{x}_{k,t})
$$
$$
\times \prod_{n'=1 \neq n}^{N} \Delta_{p_{y_{n',t}|z_{n',t}} \to \tilde{x}_{k,t}}(\tilde{x}_{k,t}). \quad (27)
$$

where $p_{\tilde{x}_{k,t}}(\tilde{x}_{k,t}) = p_{x_{k,t}}(\tilde{x}_{k,t})$ is the prior distribution of $\tilde{x}_{k,t}$ determined by the modulation of the data, and $\Delta_{\delta_k \to \tilde{x}_{k,t}}(\tilde{x}_{k,t})$ is the message provided by the prior knowledge of $\mathbf{X}_P$. The message passing process of part II can be realized by the BiG-AMP algorithm in [20]. Note that the marginal posterior $\Delta_{\tilde{x}_{k,t}}(\tilde{x}_{k,t}) = \Delta_{\delta_k \to \tilde{x}_{k,t}}(\tilde{x}_{k,t}) \Delta_{\tilde{x}_{k,t} \to \delta_k}(\tilde{x}_{k,t})$ rather than $\Delta_{\tilde{x}_{k,t} \to p_{y_{n,t}|z_{n,t}}}(\tilde{x}_{k,t})$ is needed in the BiG-AMP algorithm for complexity reduction. The discrete form of $\Delta_{\tilde{x}_{k,t}}(\tilde{x}_{k,t})$ is given by

$$
P_{\tilde{x}_{k,t}}(\tilde{x}_{k,t} = c)
$$
$$
\propto P_{\delta_k \to \tilde{x}_{k,t}}(\tilde{x}_{k,t} = c) P_{\tilde{x}_{k,t} \to \delta_k}(\tilde{x}_{k,t} = c)
$$
$$
= \frac{1}{C} \sum_{\substack{i \in \mathcal{I}_K, \omega \in \Omega \\ \text{for: } e^{j\omega} x_{i,t} = c}} P_{\delta_\Pi \to \boldsymbol{\pi}_k}(\boldsymbol{\pi}_k = \mathbf{e}_i)
$$
$$
\times \prod_{t'=1 \neq t}^{T_P} P_{\tilde{x}_{k,t'} \to \delta_k}(\tilde{x}_{k,t'} = e^{j\omega} x_{i,t'})
$$
$$
\times P_{\tilde{x}_{k,t} \to \delta_k}(\tilde{x}_{k,t} = c)
$$
$$
= \frac{1}{C} \sum_{\substack{i \in \mathcal{I}_K, \omega \in \Omega \\ \text{for } e^{j\omega} x_{i,t} = c}} P_{\delta_\Pi \to \boldsymbol{\pi}_k}(\boldsymbol{\pi}_k = \mathbf{e}_i)
$$
$$
\times \prod_{t'=1}^{T_P} P_{\tilde{x}_{k,t'} \to \delta_k}(\tilde{x}_{k,t'} = e^{j\omega} x_{i,t'}), \quad c \in \mathcal{C}. \quad (28)
$$

The estimates of $\tilde{\mathbf{H}}$ and $\tilde{\mathbf{X}}$ from the message passing iteration contain phase and permutation ambiguities. We now

describe how to estimate the phase and permutation ambiguities. Specifically, the marginal posterior of σ_k can be depicted as

$$
\Delta_{\sigma_k}(\sigma_k)
$$
$$
\propto \Delta_{\delta_k \to \sigma_k}(\sigma_k) p_{\sigma_k}(\sigma_k)
$$
$$
\propto \int_{\boldsymbol{\pi}_k, \{x_{k,t}\}_{t=1}^{T_P}} \delta(\tilde{\mathbf{x}}_{P,k} - \sigma_k \boldsymbol{\pi}_k^T \mathbf{X}_P) \prod_{t=1}^{T_P} \Delta_{\tilde{x}_{k,t} \to \delta_k}(\tilde{x}_{k,t})
$$
$$
\times \Delta_{\boldsymbol{\pi}_k \to \delta_k}(\boldsymbol{\pi}_k). \quad (29)
$$

The discrete form of the message above can be written as

$$
P_{\sigma_k}(\sigma_k = e^{j\omega})
$$
$$
= \frac{1}{C} \sum_{i \in \mathcal{I}_K} P_{\delta_\Pi \to \boldsymbol{\pi}_k}(\boldsymbol{\pi}_k = \mathbf{e}_i)
$$
$$
\times \prod_{t=1}^{T_P} P_{\tilde{x}_{k,t} \to \delta_k}(\tilde{x}_{k,t} = e^{j\omega} x_{i,t}), \quad \omega \in \Omega. \quad (30)
$$

Then, an estimate of Σ is given by $\hat{\Sigma} = \mathrm{diag}\{\hat{\sigma}_1, \hat{\sigma}_2, \ldots, \hat{\sigma}_K\}$, where $\hat{\sigma}_k = \arg\max_{\omega \in \Omega} P_{\sigma_k}(e^{j\omega})$. The marginal posterior of $\mathbf{P}$ can be depicted as

$$
\Delta_{\mathbf{P}}(\mathbf{P})
$$
$$
\propto \Delta_{\delta_\Pi \to \mathbf{P}}(\mathbf{P}) p_{\mathbf{P}}(\mathbf{P})
$$
$$
\propto \int_{\{\boldsymbol{\pi}_k\}_{k=1}^K} \delta([\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots, \boldsymbol{\pi}_K]^T - \mathbf{P}) \prod_{k=1}^{K} \Delta_{\boldsymbol{\pi}_k \to \delta_\Pi}(\boldsymbol{\pi}_k). \quad (31)
$$

We write (31) in its discrete form as

$$
P_{\mathbf{P}}(\mathbf{P} = \mathbf{P}_\ell) = \frac{1}{C} \prod_{k=1}^{K} P_{\boldsymbol{\pi}_k \to \delta_\Pi}(\boldsymbol{\pi}_k = \mathbf{p}_{\ell,k}), \quad \ell \in \mathcal{I}_{K!}. \quad (32)
$$

Then, we obtain an estimate of Π by $\hat{\Pi} = \arg\max P_{\mathbf{P}}(\mathbf{P}_\ell)$.

We are now ready to present the overall message passing algorithm in Algorithm 1. In specific, steps 1 and 2 update the estimate of the variance $\{\bar{v}_{n,t}^p(l)\}$ and the mean $\{\bar{p}_{n,t}(l)\}$ of $\{z_{n,t}\}$, where l denotes the inner iteration number. Steps 3 and 4 update the estimate of the variance $\{v_{n,t}^p(l)\}$ and the mean $\{\hat{p}_{n,t}(l)\}$ of $\{z_{n,t}\}$ by the "Onsager" correction [22]. Steps 5 and 6 give the estimate of the marginal posterior variance $\{v_{n,t}^z(l)\}$ and mean $\{\hat{z}_{n,t}(l)\}$ of $\{z_{n,t}\}$. Steps 7 and 8 calculate the inverse-residual-variances $\{v_{n,t}^s(l)\}$ and the scaled residual $\{\hat{s}_{n,t}(l)\}$. Steps 9 and 10 update the estimate of the variance $\{v_{n,k}^q(l)\}$ and mean $\{\hat{q}_{n,k}(l)\}$ of $\{\tilde{h}_{n,k}\}$ based on the messages from check nodes $\{p_{y_{n,t}|z_{n,t}}\}$. Then, steps 13 and 14 give the estimate of the marginal posterior of $\{h_{n,k}\}$ with the mean $\{\hat{h}_{n,k}(l+1)\}$ and the variance $\{v_{n,k}^h(l+1)\}$. Steps 11 and 12 process the same operations as steps 9 and 10 and steps 15 and 16 process the same operations as steps 13 and 14 for $\{\tilde{x}_{k,t} \in \mathbf{X}_D\}$. Step 17 updates the messages from $\{\tilde{x}_{k,t}\}$ to $\{\delta_k\}$, where $\mathcal{CN}(\cdot; \hat{r}_{k,t}(l), v_{k,t}^r(l))$ is a Gaussian distribution obtained by $\prod_{n=1}^{N} \Delta_{p_{n,t} \to \tilde{x}_{k,t}}(\tilde{x}_{k,t})$. Step 18 calculate the probability of $\{\boldsymbol{\pi}_k = \mathbf{e}_i\}$ with phase

Algorithm 1: The SCSE Algorithm

Input: $\mathbf{Y}$, prior distribution $p_{\tilde{\mathbf{H}}}(\tilde{\mathbf{H}})$, $p_{\tilde{\mathbf{X}}}(\tilde{\mathbf{X}})$, and $p_{\mathbf{Y}|\mathbf{Z}}(\mathbf{Y}|\mathbf{Z})$

Initialization : $\forall n, k, t, \hat{\tilde{h}}_{n,k}(1) = 0$, $v_{n,k}^{\tilde{h}} = 1$, $\hat{\tilde{x}}_{k,t}(1)$ is randomly drawn from $\mathcal{C}$, $v_{k,t}^{\tilde{x}}(1) = 1$, and $\hat{s}_{n,t}(0) = 0$

for $m = 1, \ldots, M_{max}$ %outer iteration
  for $l = 1, \ldots, L_{max}$ %inter iteration
% Message passing for part I

1: $\forall n, t: \bar{v}_{n,t}^p(l) = \sum_{k=1}^K |\hat{\tilde{h}}_{n,k}(l)|^2 v_{k,t}^{\tilde{x}}(l) + v_{n,k}^{\tilde{h}}(l)|\hat{\tilde{x}}_{k,t}(l)|^2$

2: $\forall n, t: \bar{p}_{n,t}(l) = \sum_{k=1}^K \hat{\tilde{h}}_{n,k}(l)\hat{\tilde{x}}_{k,t}(l)$

3: $\forall n, t: v_{n,t}^p(l) = \bar{v}_{n,t}^p(l) + \sum_{k=1}^K v_{n,k}^{\tilde{h}}(l)v_{k,t}^{\tilde{x}}(l)$

4: $\forall n, t: \hat{p}_{n,t}(l) = \bar{p}_{n,t}(l) - \hat{s}_{n,t}(l-1)\bar{v}_{n,t}^p(l)$

5: $\forall n, t: v_{n,t}^z(l) = \frac{v_{n,t}^p(l)\sigma^2}{v_{n,t}^p(l)+\sigma^2}$

6: $\forall n, t: \hat{z}_{n,t}(l) = \frac{v_{n,t}^p(l)\sigma^2}{v_{n,t}^p(l)+\sigma^2}(y_{n,t} - \hat{p}_{n,t}(l)) + \hat{p}_{n,t}(l)$

7: $\forall n, t: v_{n,t}^s(l) = (1 - v_{n,t}^z(l)/v_{n,t}^p(l))/v_{n,t}^p(l)$

8: $\forall n, t: \hat{s}_{n,t}(l) = (\hat{z}_{n,t}(l) - \hat{p}_{n,t}(l))/v_{n,t}^p(l)$

9: $\forall n, k: v_{n,k}^q(l) = (\sum_{t=1}^T |\hat{\tilde{x}}_{k,t}(l)|^2 v_{n,t}^s(l))^{-1}$

10: $\forall n, k: \hat{q}_{n,k}(l) = \hat{\tilde{h}}_{n,k}(l)(1 - \sum_{t=1}^T v_{k,t}^{\tilde{x}}(l)v_{n,t}^s(l)) + v_{n,k}^q(l)\sum_{t=1}^T(\hat{\tilde{x}}_{k,t}(l))^*\hat{s}_{n,t}(l)$

11: $\forall k, t: v_{k,t}^r(l) = (\sum_{n=1}^N |\hat{\tilde{h}}_{n,k}(l)|^2 v_{n,t}^s(l))^{-1}$

12: $\forall k, t: \hat{r}_{k,t}(l) = \hat{\tilde{x}}_{k,t}(l)(1 - \sum_{n=1}^N v_{n,k}^{\tilde{h}}(l)v_{k,t}^r(l)) + v_{k,t}^r(l)\sum_{n=1}^N(\hat{\tilde{h}}_{n,k}(l))^*\hat{s}_{n,t}(l)$

13: $\forall n, k: \hat{\tilde{h}}_{n,k}(l+1) = \mathbb{E}[\tilde{h}_{n,k}|\hat{q}_{n,k}(l), v_{n,k}^q(l)]$

14: $\forall n, k: v_{n,k}^{\tilde{h}}(l+1) = \mathbb{E}[|\tilde{h}_{n,k} - \hat{\tilde{h}}_{n,k}(l+1)|^2|\hat{q}_{n,k}(l), v_{n,k}^q(l)]$

15: $\forall k, t \in \mathcal{T}_D: \hat{\tilde{x}}_{k,t}(l+1) = \mathbb{E}[\tilde{x}_{k,t}|\hat{r}_{k,t}(l), v_{k,t}^r(l)]$

16: $\forall k, t \in \mathcal{T}_D: v_{k,t}^{\tilde{x}}(l+1) = \mathbb{E}[|\tilde{x}_{k,t} - \hat{\tilde{x}}_{k,t}(l+1)|^2|\hat{r}_{k,t}(l), v_{k,t}^r(l)]$
% Message passing for part II

17: $\forall k, t \in \mathcal{T}_P: P_{\tilde{x}_{k,t}\to\delta_k}^l(\tilde{x}_{k,t} = c) = \frac{1}{C}P_{\tilde{x}_{k,t}}(\tilde{x}_{k,t} = c)$
      $\times \mathcal{CN}(\tilde{x}_{k,t} = c; \hat{r}_{k,t}(l), v_{k,t}^r(l)), \quad c \in \mathcal{C}$ % Eq.(16)

18: $\forall k: p(\boldsymbol{\pi}_k = \mathbf{e}_i, \sigma_k = \omega)$
      $= \frac{1}{C}\prod_{t=1}^{T_P}\left[P_{\tilde{x}_{k,t}\to\delta_k}^l(\tilde{x}_{k,t} = e^{j\omega}x_{i,t})\right], i \in \mathcal{I}_K, \omega \in \Omega$

19: $\forall k: P_{\pi_k\to\delta_\Pi}^l(\boldsymbol{\pi}_k = \mathbf{e}_i)$
      $= \frac{1}{C}\sum_{\omega\in\Omega} p(\boldsymbol{\pi}_k = \mathbf{e}_i, \sigma_k = \omega)$ % Eq.(19),(20)

20: $\forall k: P_{\delta_\Pi\to\pi_k}^l(\boldsymbol{\pi}_k = \mathbf{e}_i) = \frac{1}{C}\sum_{\ell: \mathbf{p}_{\ell,k}=\mathbf{e}_i}^{K!}\prod_{k'=1\neq k}^K\left[P_{\pi_{k'}\to\delta_\Pi}(\boldsymbol{\pi}_{k'} = \mathbf{p}_{\ell,k'})\right], \quad i \in \mathcal{I}_K$ % Eq.(23),(24)

21: $\forall k, t \in \mathcal{T}_P: P_{\tilde{x}_{k,t}}^{l+1}(\tilde{x}_{k,t} = c) = \frac{1}{C}\sum_{i\in\mathcal{I}_K,\omega\in\Omega} P_{\delta_\Pi\to\pi_k}^l(\boldsymbol{\pi}_k = \mathbf{e}_i) p(\boldsymbol{\pi}_k = \mathbf{e}_i, \sigma_k = \omega), \quad c \in \mathcal{C}$ %$_{e^{j\omega}x_{i,t}=c}$ % Eq.(28)

22: $\forall k, t \in \mathcal{T}_P: \hat{\tilde{x}}_{k,t}(l+1) = \mathbb{E}[x_{k,t}|\hat{r}_{k,t}(l), v_{k,t}^r(l), \mathbf{X}_P]$

23: $\forall k, t \in \mathcal{T}_P: v_{k,t}^{\tilde{x}}(l+1) = \mathbb{E}[|x_{k,t} - \hat{\tilde{x}}_{k,t}(l+1)|^2|\hat{r}_{k,t}(l), v_{k,t}^r(l), \mathbf{X}_P]$

24: **if** $\sum_{n,t}|\bar{p}_{n,t}(l) - \bar{p}_{n,t}(l-1)|^2 \leq \epsilon \sum_{n,t}|\bar{p}_{n,t}(l)|^2$, **stop**
  end
  $\forall k, t: \hat{\tilde{x}}_{k,t}(1) = \hat{\tilde{x}}_{k,t}(l+1); v_{k,t}^{\tilde{x}}(l) = v_{k,t}^{\tilde{x}}(l+1);$
  $\forall n, k: \hat{\tilde{h}}_{n,k}(1) = 0; v_{n,k}^{\tilde{h}} = 1$ % Re-initialization
end
%Eliminate ambiguities

25: $\forall k: P_{\sigma_k}(\sigma_k = e^{j\omega}) = \frac{1}{C}\sum_{i\in\mathcal{I}_K} P_{\delta_\Pi\to\pi_k}^{L_{max}}(\boldsymbol{\pi}_k = \mathbf{e}_i)$
      $\times \prod_{t=1}^{T_P} P_{\tilde{x}_{k,t}\to\delta_k}^{L_{max}}(\tilde{x}_{k,t} = e^{j\omega}x_{i,t}), \quad \omega \in \Omega$ %Eq.(30)

26: $\forall k: \hat{\sigma}_k = \arg\max_{\omega\in\Omega} P_{\sigma_k}(e^{j\omega}), \hat{\Sigma} = \{\hat{\sigma}_1, \hat{\sigma}_2, \ldots, \hat{\sigma}_K\}$

27: $P_\mathbf{P}(\mathbf{P} = \mathbf{P}_\ell) = \frac{1}{C}\prod_{k=1}^K P_{\pi_k\to\delta_\Pi}^{L_{max}}(\boldsymbol{\pi}_k = \mathbf{p}_{\ell,k}), \quad \ell \in \mathcal{I}_{K!}$ %Eq.(32)

28: $\hat{\Pi} = \arg\max_{\ell\in\mathcal{I}_{K!}} P_\mathbf{P}(\mathbf{P}_\ell)$

Output: $\hat{\mathbf{H}} = \tilde{\mathbf{H}}\hat{\Sigma}\hat{\Pi}$, $\hat{\mathbf{X}}_D = \hat{\Pi}^{-1}\hat{\Sigma}^{-1}\tilde{\mathbf{X}}_D$
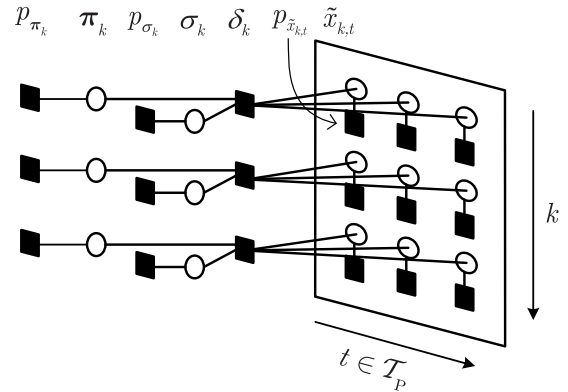

FIGURE 2. The simplified factor graph representation for Part II in Fig. 1 with $K = 3$ and $T_P = 3$.

shift $\{e^{j\omega}\}$. Steps 19 and 20 update the messages between $\boldsymbol{\pi}_k$ and Π. Step 21 updates the marginal posterior of $\{\tilde{x}_{k,t}\}$ by merging the messages from $\{p_{y_{n,t}|z_{n,t}}\}$ and $\{\delta_k\}$ with the priors $\{p_{\tilde{x}_{k,t}}(\tilde{x}_{k,t})\}$. Steps 22 and 23 give the posterior mean and variance of $\{\tilde{x}_{k,t}\} \in \mathbf{X}_P$, where the expectations are taken over the distribution given by step 19. Step 24 gives a stopping condition based on the (normalized) change of $\bar{p}_{n,t}(l)$ in two consecutive iteration and a user-defined parameter ϵ. Steps 25 and 27 give the posterior message of $\{\sigma_k\}$ and $\{\boldsymbol{\pi}_k\}$. Steps 26 and 28 give estimates of Σ and Π. Note that steps 25-28 are out of the loop since the estimates of the phase and permutation ambiguities are not needed in the iterative process but are needed in correcting the estimates of $\mathbf{H}$ and $\mathbf{X}$ after message passing.

It is interesting to compare the joint channel-and-signal estimation scheme in [9] with SCSE in Algorithm 1. In fact, Algorithm 1 can be modified for the joint channel-and-signal scheme as follows: i) In initialization, set $\hat{\tilde{x}}_{k,t}(1) = x_{k,t}$, $\forall k, t \in \mathcal{T}_P$; ii) delete steps 17-23, and 25-28. That is, the main difference of SCSE from the joint channel-and-signal estimation scheme is that the former includes the estimation of the phase and permutation ambiguities, i.e., Σ, and Π, in the iterative message passing process. We will show by numerical simulations that SCSE is able to significantly outperform the joint channel-and-signal estimation scheme in the short-pilot regime.

C. S-SCSE ALGORITHM

The SCSE algorithm is computationally infeasible for a relatively large k, since it involves enumeration over all length-k permutations in step 20 of Algorithm 1. To reduce the complexity, we relax the constraint that Π is a permutation to the one that each row k of Π (denoted by $\boldsymbol{\pi}_k^T$) is independently taken from the set $\{\mathbf{e}_\ell\}_{\ell=1}^K$. That is,

$$p_\Pi(\Pi) = p_\Pi(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots, \boldsymbol{\pi}_K) \approx \prod_{k=1}^K p_{\boldsymbol{\pi}_k}(\boldsymbol{\pi}_k) \quad (33)$$

where $p_{\boldsymbol{\pi}_k}(\boldsymbol{\pi}_k) = 1/K$, and $\boldsymbol{\pi}_k \in \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_K\}$. The corresponding factor graph of part II in Fig. 1 is given in Fig. 2.

Compared with part II in Fig. 1, the factor graph in Fig. 2 is almost the same, except that the nodes $\{p_\mathbf{P}, \mathbf{P}, \sigma_\Pi\}$ are replaced by $\{p_{\boldsymbol{\pi}_k}\}$.

We now describe message passing over the factor graph in Fig. 2. The messages from $\tilde{x}_{k,t}$ to δ_k and from σ_k to δ_k have the same form as in (16) and (17). The message from $\boldsymbol{\pi}_k$ to δ_k is given by

$$\Delta_{\boldsymbol{\pi}_k \to \delta_k}(\boldsymbol{\pi}_k) = p_{\boldsymbol{\pi}_k}(\boldsymbol{\pi}_k) = \frac{1}{K} \sum_{i=1}^{K} \delta(\boldsymbol{\pi}_k = \mathbf{e}_i). \quad (34)$$

Then, the message from δ_k to $\tilde{x}_{k,t}$ is given by

$$\Delta_{\delta_k \to \tilde{x}_{k,t}}(\tilde{x}_{k,t})$$
$$\propto \int_{\sigma_k, \boldsymbol{\pi}_k, \{x_{k,t'}\}_{t'=1 \neq t}^{T_P}} \delta(\tilde{\mathbf{x}}_{P,k} - \sigma_k \boldsymbol{\pi}_k^T \mathbf{X}_P)$$
$$\times p_{\boldsymbol{\pi}_k}(\boldsymbol{\pi}_k) p_{\sigma_k}(\sigma_k) \prod_{t'=1 \neq t}^{T_P} \Delta_{\tilde{x}_{k,t'} \to \delta_k}(\tilde{x}_{k,t'}). \quad (35)$$

Since $\tilde{x}_{k,t}$, $\boldsymbol{\pi}_k$, and σ_k are discrete variables, we can write the above message in its discrete form as

$$P_{\delta_k \to \tilde{x}_{k,t}}(\tilde{x}_{k,t} = c)$$
$$= \frac{1}{C} \sum_{\substack{i \in \mathcal{I}_K, \omega \in \Omega \\ \text{for: } e^{j\omega} x_{i,t} = c}} \prod_{t'=1 \neq t}^{T_P} P_{\tilde{x}_{k,t'} \to \delta_k}(\tilde{x}_{k,t'} = e^{j\omega} x_{i,t'}), \quad c \in \mathcal{C}.$$
$$(36)$$

Then, for any $c \in \mathcal{C}$, the marginal posterior of $\tilde{x}_{k,t}$ can be updated as

$$P_{\tilde{x}_{k,t}}(\tilde{x}_{k,t} = c)$$
$$\propto P_{\delta_k \to \tilde{x}_{k,t}}(\tilde{x}_{k,t} = c) P_{\tilde{x}_{k,t} \to \delta_k}(\tilde{x}_{k,t} = c)$$
$$= \frac{1}{C} \sum_{\substack{i \in \mathcal{I}_K, \omega \in \Omega \\ \text{for: } e^{j\omega} x_{i,t} = c}} \prod_{t'=1}^{T_P} P_{\tilde{x}_{k,t'} \to \delta_k}(\tilde{x}_{k,t'} = e^{j\omega} x_{i,t'}), \quad c \in \mathcal{C}.$$
$$(37)$$

The other messages are calculated by following the SCSE algorithm. Compared to SCSE, S-SCSE omits the calculation of $P_{\delta_\Pi \to \boldsymbol{\pi}_k}(\boldsymbol{\pi}_k = \mathbf{e}_i)$ in (23), which significantly reduces the computation complexity from $\mathcal{O}(K!)$ to $\mathcal{O}(K^2)$.

The phase and permutation ambiguities are estimated as follows. The marginal posterior of σ_k can be depicted as

$$\Delta_{\sigma_k}(\sigma_k)$$
$$\propto \Delta_{\delta_k \to \sigma_k}(\sigma_k) p_{\sigma_k}(\sigma_k)$$
$$\propto \int_{\boldsymbol{\pi}_k, \{x_{k,t}\}_{t=1}^{T_P}} \delta(\tilde{\mathbf{x}}_{P,k} - \sigma_k \boldsymbol{\pi}_k^T \mathbf{X}_P) \prod_{t=1}^{T_P} \Delta_{\tilde{x}_{k,t} \to \delta_k}(\tilde{x}_{k,t}). \quad (38)$$

The discrete form of the above message can be written as

$$P_{\sigma_k}(\sigma_k = e^{j\omega})$$
$$= \frac{1}{C} \sum_{i \in \mathcal{I}_K} \prod_{t=1}^{T_P} P_{\tilde{x}_{k,t} \to \delta_k}(\tilde{x}_{k,t} = e^{j\omega} x_{i,t}), \quad \omega \in \Omega. \quad (39)$$

Algorithm 2: The S-SCSE Algorithm

Input: $\mathbf{Y}$, prior distribution $p_{\tilde{\mathbf{H}}}(\tilde{\mathbf{H}})$, $p_{\tilde{\mathbf{X}}}(\tilde{\mathbf{X}})$, and $p_{\mathbf{Y}|\mathbf{Z}}(\mathbf{Y}|\mathbf{Z})$

Initialization : $\forall n, k, t, \hat{\tilde{h}}_{n,k}(1) = 0, v_{n,k}^{\tilde{h}} = 1, \hat{\tilde{x}}_{k,t}(1)$ is randomly drawn from $\mathcal{C}, v_{k,t}^{\tilde{x}}(1) = 1$, and $\hat{s}_{n,t}(0) = 0$

for $m = 1, \ldots, M_{max}$ %outer iteration
  for $l = 1, \ldots, L_{max}$ %inter iteration
    % Message passing for part I
    Perform steps 1-16 in Algorithm 1.
    % Message passing for part II
    Perform steps 17-18 in Algorithm 1.
1: $\forall (k, t) \in \mathbf{X}_P$: $P_{\tilde{x}_{k,t}}^{l+1}(\tilde{x}_{k,t} = c) = \frac{1}{C}$
    $\times \sum_{\substack{i \in \mathcal{I}_K, \omega \in \Omega \\ e^{j\omega} x_{i,t} = c}} p(\boldsymbol{\pi}_k = \mathbf{e}_i, \sigma_k = \omega)$ $c \in \mathcal{C}$
                    % Eq. (37)
2: $\forall k, t \in \mathcal{T}_P$: $\hat{\tilde{x}}_{k,t}(l+1) = \mathbb{E}[x_{k,t} | \hat{r}_{k,t}(l), v_{k,t}^r(l), \mathbf{X}_P]$
3: $\forall k, t \in \mathcal{T}_P$: $v_{k,t}^{\tilde{x}}(l+1) =$
      $\mathbb{E}[|x_{k,t} - \hat{\tilde{x}}_{k,t}(l+1)|^2 | \hat{r}_{k,t}(l), v_{k,t}^r(l), \mathbf{X}_P]$
4: **if** $\sum_{n,t} |\bar{p}_{n,t}(l) - \bar{p}_{n,t}(l-1)|^2 \leq \epsilon \sum_{n,t} |\bar{p}_{n,t}(l)|^2$, **stop**
  end
  $\forall k, t$: $\hat{\tilde{x}}_{k,t}(1) = \hat{\tilde{x}}_{k,t}(l+1); v_{k,t}^{\tilde{x}}(l) = v_{k,t}^{\tilde{x}}(l+1);$
  $\forall n, k$: $\hat{\tilde{h}}_{n,k}(1) = 0; v_{n,k}^{\tilde{h}} = 1$ % Re-initialization
end
% Eliminate ambiguities
5: $\forall k$: $P_{\sigma_k}(\sigma_k = e^{j\omega})$
    $= \frac{1}{C} \sum_{i \in \mathcal{I}_K} \prod_{t=1}^{T_P} P_{\tilde{x}_{k,t} \to \delta_k}^{L_{max}}(\tilde{x}_{k,t} = e^{j\omega} x_{i,t}), \quad \omega \in \Omega$
                  % Eq. (39)
6: $\forall k$: $P_{\boldsymbol{\pi}_k}(\boldsymbol{\pi}_k = \mathbf{e}_\ell)$
    $= \frac{1}{C} \sum_{\omega \in \Omega} \prod_{t=1}^{T_P} P_{\tilde{x}_{k,t} \to \delta_k}^{L_{max}}(\tilde{x}_{k,t} = e^{j\omega} x_{i,t}), \quad \ell \in \mathcal{I}_K$
                  % Eq. (41)
7: $\forall k$: $\hat{\sigma}_k = \arg \max_{\omega \in \Omega} P_{\sigma_k}(e^{j\omega})$, $\hat{\Sigma} = \{\hat{\sigma}_1, \hat{\sigma}_2, \ldots, \hat{\sigma}_K\}$
8: $\forall k$: $\hat{\boldsymbol{\pi}}_k = \arg \max_{\ell \in \mathcal{I}_K} P_{\boldsymbol{\pi}_k}(\mathbf{e}_\ell)$, $\hat{\Pi} = \{\hat{\boldsymbol{\pi}}_1, \hat{\boldsymbol{\pi}}_2, \ldots, \hat{\boldsymbol{\pi}}_K\}$
Output: : $\hat{\mathbf{H}} = \tilde{\mathbf{H}} \hat{\Sigma} \hat{\Pi}, \hat{\mathbf{X}}_D = \hat{\Pi}^{-1} \hat{\Sigma}^{-1} \tilde{\mathbf{X}}_D$

Then, an estimate of Σ is given by $\hat{\Sigma} = \text{diag}\{\hat{\sigma}_1, \hat{\sigma}_2, \ldots, \hat{\sigma}_K\}$, where $\hat{\sigma}_k = \arg \max_{\omega \in \Omega} P_{\sigma_k}(e^{j\omega})$. The marginal posterior of $\boldsymbol{\pi}_k$ can be depicted as

$$\Delta_{\boldsymbol{\pi}_k}(\boldsymbol{\pi}_k)$$
$$\propto \Delta_{\delta_k \to \boldsymbol{\pi}_k}(\boldsymbol{\pi}_k) p_{\boldsymbol{\pi}_k}(\boldsymbol{\pi}_k)$$
$$\propto \int_{\sigma_k, \{x_{k,t}\}_{t=1}^{T_P}} \delta(\tilde{\mathbf{x}}_{P,k} - \sigma_k \boldsymbol{\pi}_k^T \mathbf{X}_P) \prod_{t=1}^{T_P} \Delta_{\tilde{x}_{k,t} \to \delta_k}(\tilde{x}_{k,t}). \quad (40)$$

Similarly, the discrete form of the above message can be written as

$$P_{\boldsymbol{\pi}_k}(\boldsymbol{\pi}_k = \mathbf{e}_i)$$
$$= \frac{1}{C} \sum_{\omega \in \Omega} \prod_{t=1}^{T_P} P_{\tilde{x}_{k,t} \to \delta_k}(\tilde{x}_{k,t} = e^{j\omega} x_{i,t}), \quad i \in \mathcal{I}_K. \quad (41)$$

Then, an estimate of Π is given by $\hat{\Pi} = [\hat{\boldsymbol{\pi}}_1, \hat{\boldsymbol{\pi}}_2, \ldots, \hat{\boldsymbol{\pi}}_K]^T$, where $\hat{\boldsymbol{\pi}}_k = \arg \max_{\ell \in \mathcal{I}_K} P_{\boldsymbol{\pi}_k}(\mathbf{e}_\ell)$.

The S-SCSE algorithm is presented in Algorithm 2. In Algorithm 2, the S-SCSE algorithm performs the

steps 1-18 of Algorithm 1 first. Then, step 1 of Algorithm 2 updates the marginal posterior of $\{\tilde{x}_{k,t}\}$ by S-SCSE algorithm. Steps 2 and 3 give the posterior mean and variance of $\{\tilde{x}_{k,t}\} \in \mathbf{X}_{\mathrm{P}}$, where the expectations are taken over the distribution given by steps 1. Steps 5 and 6 give the posterior messages of σ_k and $\boldsymbol{\pi}_k$ by S-SCSE algorithm, respectively. Steps 7 and 8 give estimates of Σ and Π, respectively.[3] In addition, the damping technique is used in the algorithm to improve convergence in simulation. We refer readers to [20] and [23] for more details.

D. PARAMETER TUNING

Note that both the SCSE and S-SCSE algorithms require the knowledge of the distributions of $p_{\tilde{\mathbf{H}}}(\tilde{\mathbf{H}})$, $p_{\tilde{\mathbf{X}}}(\tilde{\mathbf{X}})$, and $p_{\mathbf{Y}|\mathbf{Z}}(\mathbf{Y}|\mathbf{Z})$ to perform the message passing process. However, the parameters $\{\rho, \sigma_{h,k}^2, \forall k, N_0\}$ are usually difficult to acquire prior to the detection procedure [20]. Therefore, these model parameters need to be estimated as well.

In this paper, we use the EM algorithm to tune these model parameters by taking $\tilde{\mathbf{H}}$ and $\tilde{\mathbf{X}}$ as the hidden variables. The specific update rules are given as follows.

$$\rho^{(m+1)} = \arg\max_{\rho} \mathbb{E}\left[\log p\left(\tilde{\mathbf{H}}, \tilde{\mathbf{X}}, \mathbf{Y}; \rho, (\sigma_{h,1}^2)^{(m)}, \ldots,\right.\right.$$
$$\left.\left. (\sigma_{h,K}^2)^{(m)}, N_0^{(m)}\right)\right]. \quad (42)$$

$$(\sigma_{h,k}^2)^{(m+1)} = \arg\max_{\sigma_{h,k}^2} \mathbb{E}\left[\log p\left(\tilde{\mathbf{H}}, \tilde{\mathbf{X}}, \mathbf{Y}; \rho^{(m+1)},\right.\right.$$
$$(\sigma_{h,1}^2)^{(m+1)}, \ldots, (\sigma_{h,k-1}^2)^{(m+1)}, (\sigma_{h,k}^2),$$
$$\left.\left. (\sigma_{h,k+1}^2)^{(m)}, \ldots (\sigma_{h,K}^2)^{(m)}, N_0^{(m)}\right)\right], \quad \forall k.$$
$$(43)$$

$$N_0^{(m+1)} = \arg\max_{N_0} \mathbb{E}\left[\log p\left(\tilde{\mathbf{H}}, \tilde{\mathbf{X}}, \mathbf{Y}; \rho^{(m+1)},\right.\right.$$
$$\left.\left. (\sigma_{h,1}^2)^{(m+1)}, \ldots, (\sigma_{h,K}^2)^{(m+1)}, N_0\right)\right]. \quad (44)$$

The EM update is performed in each outer iteration and the expectations in (42)-(44) are taken over the approximate marginal posteriors $\left\{p_{\tilde{h}_{n,k}|\mathbf{Y}}, p_{\tilde{x}_{k,t}|\mathbf{Y}}, p_{\tilde{z}_{n,t}|\mathbf{Y}}\right\}_{\forall n,k,t}$ obtained from every L_{max}th inner iteration.

E. COMPLEXITY ANALYSIS

We now compare the computational complexity of our proposed algorithms with the existing approaches. Since both the joint channel-and-signal (JCSE) scheme in [9] and the blind detection scheme in [18] are based on the BiG-AMP algorithm [20], we only need to compare the computational complexity of the BiG-AMP, SCSE, and S-SCSE algorithms.

The computational complexity in steps 4-8 of Algorithm 1 is $\mathcal{O}(NT)$, and that in steps 1-3 and steps 9-16 is $\mathcal{O}(NK+KT)$.

---

[3]To identify users uniquely, T_{P} is required to be large enough to ensure that for each $\mathbf{x}_{\mathrm{P},k}$ and $\mathbf{x}_{\mathrm{P},k'}$, $\mathbf{x}_{\mathrm{P},k} \neq e^{j\omega}\mathbf{x}_{\mathrm{P},k'}$, $\omega \in \Omega$. This implies that, when the data are modulated by quadrature phase shift keying, T_{P} should be no less than $1 + \lceil \frac{1}{2}\log_2 K \rceil$, where one symbol is used to correct the phase shift of each user k, and $\lceil \frac{1}{2}\log_2 K \rceil$ symbols are used to guarantee that the pilot sequences of the K users are different from each other.

TABLE 1. Computational complexity.

Method	Complexity
JCSE	$\mathcal{O}(NT) + \mathcal{O}(NK + KT)$
Blind detection	$\mathcal{O}(NT) + \mathcal{O}(NK + KT)$
SCSE	$\mathcal{O}(NT) + \mathcal{O}(NK + KT) + \mathcal{O}(K^2 K!)$
S-SCSE	$\mathcal{O}(NT) + \mathcal{O}(NK + KT) + \mathcal{O}(K^2 T_{\mathrm{P}})$

Since the BiG-AMP algorithm only perform steps 1-16 in Algorithm 1 for each iteration, the computational complexity of the BiG-AMP algorithm is $\mathcal{O}(NT)+\mathcal{O}(NK+KT)$ [20]. The computational complexity of steps 17-21 in Algorithm 1 is $\mathcal{O}(KT_{\mathrm{P}})$, $\mathcal{O}(K^2 T_{\mathrm{P}})$, $\mathcal{O}(K)$, $\mathcal{O}(K^2 K!)$, and $\mathcal{O}(K^2 T_{\mathrm{P}})$, respectively. With the increase of K, $\mathcal{O}(K^2 K!)$ dominates the complexity. Then, the overall computational complexity of the SCSE algorithm is $\mathcal{O}(NT)+\mathcal{O}(NK+KT)+\mathcal{O}(K^2 K!)$ per iteration. The computational complexity of step 1 in Algorithm 2 is $\mathcal{O}(K^2 T_{\mathrm{P}})$. Thus, the computational complexity of the S-SCSE algorithm is $\mathcal{O}(NT)+\mathcal{O}(NK+KT)+\mathcal{O}(K^2 T_{\mathrm{P}})$ per iteration. The computational complexities for our considered algorithms are summarized in Table 1.

From Table 1, we see that, compared to the JCSE and blind detection schemes, the complexity of the S-SCSE algorithm is dominated by the third term $\mathcal{O}(K^2 K!)$ caused by the estimation of Σ and Π. By relaxing the permutation constraint, the S-SCSE algorithm can significantly reduce the computational complexity of estimating Σ and Π from $\mathcal{O}(K^2 K!)$ to $\mathcal{O}(K^2 T_{\mathrm{P}})$.

F. METRIC FOR RANDOM INITIALIZATIONS

The semi-blind detection problem in (12) is non-convex, and the SCSE and S-SCSE algorithms are prone to be stuck at local optima. To alleviate this issue, multiple random initializations and multiple re-initializations are conducted.

We next describe how to choose a desirable result among multiple random initializations. In a practical receiver, the metrics such as the mean-square error of the channel and the symbol error rate of the signal are not useful in evaluating the performance of random initializations since the ground truth is not available to the receiver. In this regard, we propose to use the following heuristic metric for evaluating random initializations:

$$J(\tau) = \|\mathbf{Y} - \mathbf{H}(\tau)\mathbf{X}(\tau)\|_{\mathrm{F}}^2 \quad (45)$$

where τ is the index of random initializations. We choose the initialization with the minimum value of $J(\tau)$.

V. NUMERICAL RESULTS

In simulations, the signals are taken from quadrature phase shift keying (QPSK) or 16 QAM with Gray-mapping. We set $P_k = 1, \forall k$. The SNR is defined by $\frac{K}{N_0}$. Following [9], [18], we divide all the AoA into N grids and assume that all received signals from angle $e^{-j2\pi\frac{(n-1)d}{2\lambda}\cos(\theta_{\ell,k})}$ to angle $e^{-j2\pi\frac{nd}{2\lambda}\cos(\theta_{\ell,k})}$ belong to the nth grid. Thus, the array steering matrix $\mathbf{A}_r$ can be regarded as a DFT matrix. The aggregated channel gains are generated from the B-CSCG distribution
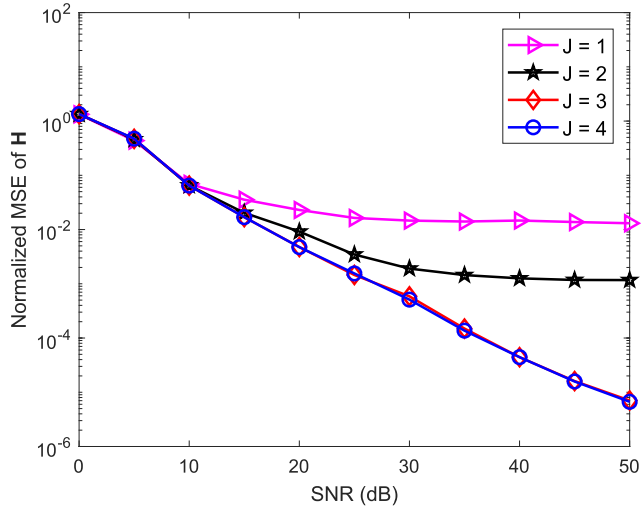
FIGURE 3. The normalized MSE of H versus SNR with the number of initializations ranging from 1 to 4 for the S-SCSE scheme with QPSK modulation. $K = 20$, $N = 200$, $\rho = 0.2$, $T = 50$, $T_P = 4$, and $\sigma_{h,min}^2 = 1$.

FIGURE 4. Comparison of the BER of X versus the number of users K for SCSE and S-SCSE with the number of pilots $T_P = 3$, 4, and 5, SNR = 0 dB, $N/K = 10$, $\rho = 0.5$, $T = 50$, and $\sigma_{h,min}^2 = 1$.

FIGURE 5. Comparison of the normalized MSE of H versus SNR for the OMP, CoSaMP, Turbo-CS, JCSE, BD, S-SCSE, and the LB-X with the pilots number $T_P = 8$, $K = 20$, $N = 128$, $\rho = 0.3$, $T = 50$, and $\sigma_{h,min}^2 = 1$.

in (4). The channel powers $\sigma_{h,k}^2$, $\forall k$ are randomly drawn from a uniform distribution over $[\sigma_{h,min}^2, 1]$. For the simulated algorithms, the maximum number of inner iterations L_{max} is set to 200, and the maximum number of outer iterations M_{max} is set to 10. The simulation results presented in this paper are obtained by taking average over 100 random realizations. We compare the numerical results of various approaches, as listed below.

- OMP: A separate channel-and-signal detection approach to estimate **H** by orthogonal matching pursuit (OMP) [24] with the pilots only.
- CoSaMP: A separate channel-and-signal detection approach to estimate **H** by compressive sampling matching pursuit (CoSaMP) [25] with the pilots only.
- Turbo-CS: A separate channel-and-signal detection approach to estimate **H** by turbo compressed sensing (Turbo-CS) [26] with the pilots only.
- JCSE: The joint channel-and-signal detection scheme in [9].
- BD: The blind detection scheme based on the BiG-AMP algorithm [20].
- SCSE: The SCSE algorithm proposed in this paper.
- S-SCSE: The S-SCSE algorithm proposed in this paper.
- LB-**H**: To estimate **H** by the BiG-AMP algorithm in [20] with perfectly known **X**.
- LB-**X**: To estimate **X** by the BiG-AMP algorithm in [20] with perfectly known positions of the non-zero elements of **H**.

Fig. 3 compares the normalized mean-square error (MSE) of **H** versus SNR with different numbers of random initializations varying from $J = 1$ to $J = 4$ for the S-SCSE algorithm with QPSK modulation. The other settings are $K = 20$, $N = 200$, $\rho = 0.2$, $T = 50$, and $\sigma_{h,min}^2 = 1$. We see that with the metric in (45), random initialization substantially improve the performance of the semi-blind scheme.
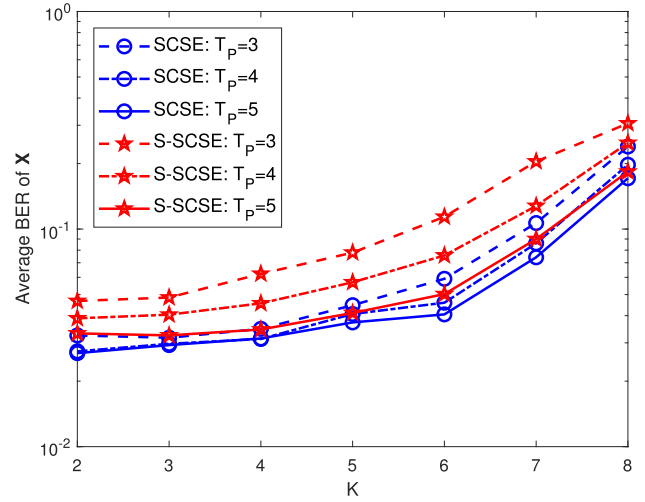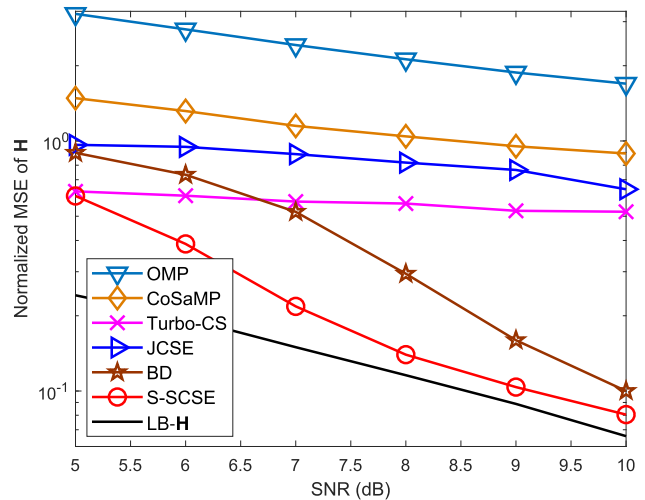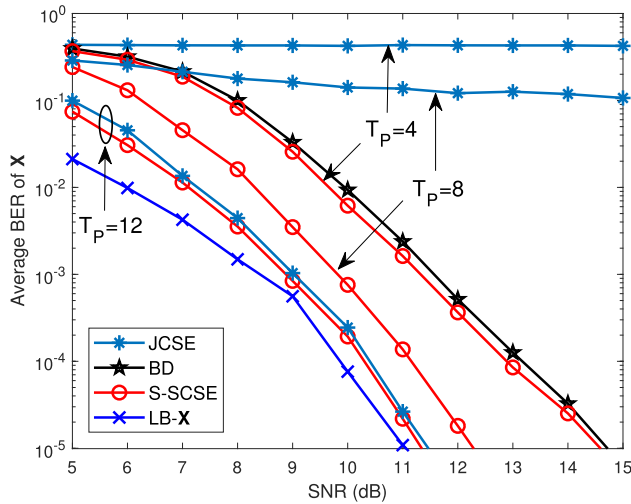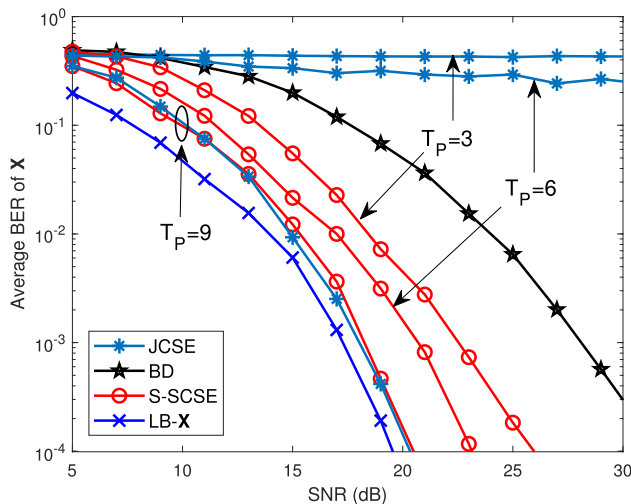
Fig. 4 compares the average bit error rate (BER) of **X** versus the number of users K for SCSE and S-SCSE with the number of pilots $T_P = 3$, 4, and 5. The other settings are SNR = 0 dB, $N/K = 10$, $\rho = 0.5$, $T = 50$, $\sigma_{h,min}^2 = 1$, and $J = 5$. We can see that for a relatively large T_P (say, $T_P = 5$ for the configuration in Fig. 4), S-SCSE is able to perform close to SCSE. Note that due to high computational complexity for SCSE, we henceforth only present the simulation results of S-SCSE.

Fig. 5 compares the normalized mean-square error (MSE) of **H** versus SNR for the OMP, CoSaMP, Turbo-CS, JCSE, BD, S-SCSE, and LB-**X** with the pilots number $T_P = 8$. The other setting are $K = 20$, $N = 128$, $\rho = 0.3$, $T = 50$, $\sigma_{h,min}^2 = 1$, and $J = 5$. From Fig. 5, we see that our S-SCSE algorithm significantly outperforms the training-based schemes (including OMP, CoSaMP, Turbo-CS, and JCSE) and blind detection scheme. We also see that with the

(a) QPSK

(b) 16QAM

FIGURE 6. The BER of X versus SNR for the JCSE, BD, S-SCSE, and LB-X with different numbers of pilots. $K = 20$, $N = 128$, $\rho = 0.3$, $T = 50$, and $\sigma^2_{h,min} = 1$.

FIGURE 7. Comparison of the BER of X versus SNR for no tuning, EM, and known $\sigma_{h,k}$ with $-10\log_{10}(\sigma^2_{h,min}) = 10/15/20$ dB of the S-SCSE algorithm, $K = 20$, $N = 128$, $\rho = 0.3$, and $T = 50$.

In Fig. 6(a), we see that for $T_P = 4$ and 8, S-SCSE significantly outperforms the JCSE scheme. We also see that for $T_P = 4$, S-SCSE slightly outperforms the blind detection scheme, while for $T_P = 8$, SCSE outperforms the blind detection scheme by about 4 dB at BER = 10^{-5}. For $T_P = 12$, the S-SCSE and JCSE schemes perform close to each other and approximate the lower bound. The reason is that in this case T_P is large enough to provide a relatively accurate initial channel estimate, and so the training-based scheme can work well. In Fig. 6(b), S-SCSE can outperforms the blind detection scheme by about 6 dB at BER = 10^{-3} for $T_P = 3$, and by about 10 dB at BER = 10^{-4} for $T_P = 9$. Similarly, S-SCSE significantly outperforms the JCSE scheme for $T_P = 3$ and 6. The S-SCSE and JCSE schemes perform close to each other for $T_P = 12$. Fig. 6 shows that both the JCSE and S-SCSE schemes achieve better performance with the increase of T_P. However, such performance improvement is achieved at the cost of a decrease in spectrum efficiency, since the pilots cannot transmit information. This issue will be elaborated later in Fig. 9.

We next study the impact of large-scale fading on the system performance. In simulations, we set $-10\log_{10}(\sigma^2_{h,min}) = 10/15/20$ dB. The other setting are $K = 20$, $N = 128$, $\rho = 0.3$, $T = 50$, and $J = 5$. We consider the different configurations of the S-SCSE algorithm: i) no tuning (in which $\sigma^2_{h,k} = 1, \forall k$); ii) EM (in which the EM algorithm in Section IV-D is used for learning $\{\sigma^2_{h,k}\}$); iii) lower bound (in which $\{\sigma^2_{h,k}\}$ are exactly known by the receiver in prior). From Fig. 7, we see that all the three approach perform close to each other when $-10\log_{10}(\sigma^2_{h,min}) = 10$ dB, whereas the EM approach significantly outperforms the no tuning approach and performs close to the lower bound for $-10\log_{10}(\sigma^2_{h,min}) = 15$ and 20 dB. This implies that the S-SCSE algorithm with EM tuning is able to efficiently handle the effect of large-scale fading.

increase of SNR, the performance of our S-SCSE algorithm can approach that of the LB-X.

Fig. 6 presents the bit error rate (BER) of X versus SNR for the JCSE, BD, S-SCSE, and LB-X with different numbers of pilot symbols. The other settings are $K = 20$, $N = 128$, $\rho = 0.3$, $T = 50$, and $\sigma^2_{h,min} = 1$. For Fig. 6(a), we set $J = 5$, while for Fig. 6(b), we set $J = 12$. Note that when the modulation is changed from QPSK (in Fig. 6(a)) to 16-QAM (in Fig. 6(b)), a lager number of random initializations is required to ensure stable semi-blind detection. Also note that the blind detection system needs one reference symbol and a user label. For the simulation settings considered here, this amounts to a cost of $1 + \lceil \frac{1}{2} \log_2 K \rceil = 4$ symbols for QPSK modulation, and $1 + \lceil \frac{1}{2} \log_4 K \rceil = 3$ symbols for 16-QAM modulation.
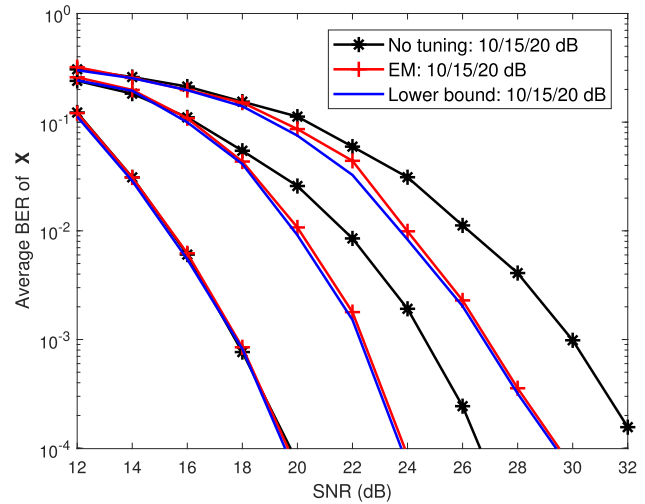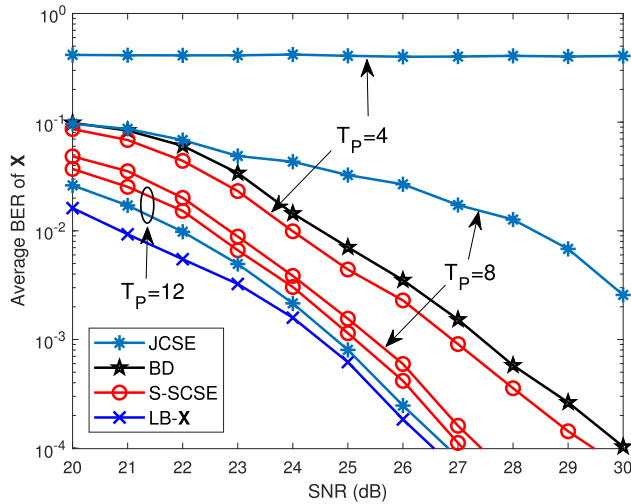
FIGURE 8. The BER of X versus SNR for the JCSE, BD, and S-SCSE with different numbers of pilots. $K = 20$, $N = 128$, $\rho = 0.3$, $T = 50$, and $-10\log_{10}(\sigma_{h,min}^2) = 20$ dB.
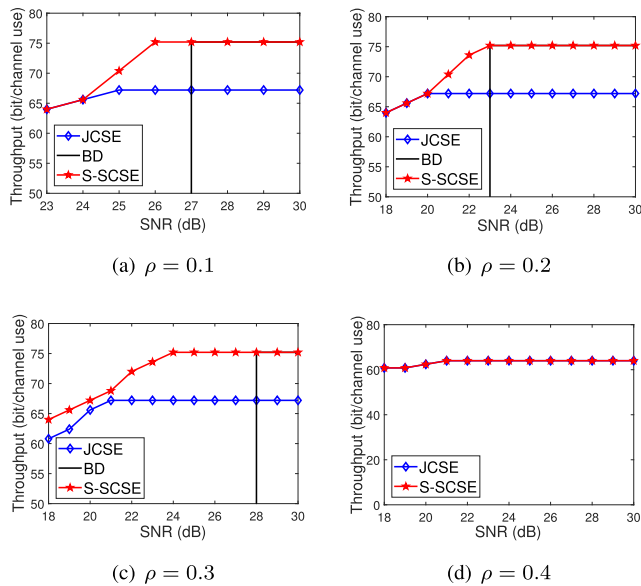

FIGURE 9. Comparison of the throughput between different sparsity with 16 QAM modulation versus SNR for the JCSE, BD, and S-SCSE, $K = 20$, $N = 200$, $T = 50$, and $\sigma_{h,min}^2 = 1$.

We now compare the performance of the various schemes in the presence of large-scale fading. In simulations, we set $-10\log_{10}(\sigma_{h,min}^2) = 20$ dB, and the EM algorithm in Section IV-D is employed for the tuning of $\{\sigma_{h,k}^2\}$. The other settings are the same as those in Fig. 6(a). From Fig. 8, we see that the trends of the curves are very similar to those in Fig. 6(a), except that the SNR is shifted by about 15 dB.

Fig. 9 shows the throughput of the JCSE, BD, and S-SCSE with 16-QAM and Gray mapping versus SNR. We say that a system performs successful recovery when BER $< 10^{-3}$. For the S-SCSE and JCSE schemes, for each given SNR, we increase the number of pilots T_P until the system performs successful recovery. For BD, T_P is fixed at 3. Then, the throughput is calculated by $4K(1 - T_P/T)$ bit per channel use. The other settings are $K = 20$, $N = 200$,

$T = 50$, and $\sigma_{h,min}^2 = 1$. For example, in the third subfigure of $\rho = 0.3$, when SNR $= 24$ dB, the required pilots to ensure successfully recovery for S-SCSE and JCSE are $T_P = 3$ and $T_P = 9$, respectively. So the throughput of the two schemes are calculated by $4 \times 20 \times (1 - 3/50) = 75.2$ and $4 \times 20 \times (1 - 9/50) = 65.6$, respectively. From Fig. 9, we see that S-SCSE considerably outperforms the JCSE scheme for $\rho = 0.1, 0.2$, and 0.3. For $\rho = 0.4$, the sparsity level ρ is too large so that the blind matrix factorization cannot provide much useful information. Both the S-SCSE and JCSE schemes rely on the knowledge of pilots for channel-and-signal estimation, and perform closely in Fig. 9(d). For comparison, we also include the SNR threshold beyond which the blind detection scheme is able to perform successful recovery. Note that the threshold is not included in Fig. 9(d) since for $\rho = 0.4$, the blind detection scheme does not work in the SNR range of interest. We see that the blind detection scheme works well only when the SNR is sufficiently high. This demonstrates the advantage of semi-blind detection.

VI. CONCLUSIONS
In this paper, we proposed a semi-blind signal detection scheme for uplink massive MIMO in which short pilot sequences are inserted into user packets and the knowledge of pilots is intergraded into the message passing algorithms for reliable matrix factorization. We derived two semi-blind estimation algorithms, namely SCSE and S-SCSE, based on the message-passing principles. In specific, the S-SCSE is a simplified version of SCSE with much lower computation complexity, but achieve almost the same performance. Numerical results showed that our proposed semi-blind scheme substantially outperforms the existing blind detection and training-based schemes in the short-pilot regime.

ACKNOWLEDGMENT

REFERENCES
[1] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[2] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[3] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[4] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[5] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[6] T. L. Marzetta, "How much training is required for multiuser MIMO?" in *Proc. 40th Asilomar Conf. Signals, Syst. Comput.*, Oct./Nov. 2006, pp. 18–22.

[7] M. Coldrey and P. Bohlin, ''Training-based MIMO systems—Part I: Performance comparison,'' *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5464–5476, Nov. 2007.

[8] X. Yuan, C. Fan, and Y. Zhang, ''Fundamental limits of training-based uplink multiuser MIMO systems,'' *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7544–7558, Nov. 2018.

[9] C. K. Wen, C. J. Wang, S. Jin, K. K. Wong, and P. Ting, ''Bayes-optimal joint channel-and-data estimation for massive MIMO with low-precision ADCs,'' *IEEE Trans. Signal Process.*, vol. 64, no. 10, pp. 2541–2556, May 2016.

[10] B. Hassibi and B. M. Hochwald, ''How much training is needed in multiple-antenna wireless links?'' *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[11] M. Honig, U. Madhow, and S. Verdu, ''Blind adaptive multiuser detection,'' *IEEE Trans. Inf. Theory*, vol. 41, no. 4, pp. 944–960, Jul. 1995.

[12] X. Wang and H. V. Poor, ''Blind multiuser detection: A subspace approach,'' *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 677–690, Mar. 1998.

[13] L. Zheng and D. N. C. Tse, ''Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels,'' *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[14] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, ''Compressed channel sensing: A new approach to estimating sparse multipath channels,'' *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, Jun. 2010.

[15] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, ''A coordinated approach to channel estimation in large-scale multiple-antenna systems,'' *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, Mar. 2013.

[16] R. R. Müller, L. Cottatellucci, and M. Vehkaperä, ''Blind pilot decontamination,'' *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 773–786, Mar. 2014.

[17] M. Masood, L. H. Afify, and T. Y. Al-Naffouri, ''Efficient coordinated recovery of sparse channels in massive MIMO,'' *IEEE Trans. Signal Process.*, vol. 63, no. 1, pp. 104–118, Jan. 2014.

[18] J. Zhang, X. Yuan, and Y. Zhang, ''Blind signal detection in massive MIMO: Exploiting the channel sparsity,'' *IEEE Trans. Wireless Commun.*, vol. 66, no. 2, pp. 700–712, Feb. 2018.

[19] J. P. Vila and P. Schniter, ''Expectation-maximization Gaussian-mixture approximate message passing,'' *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4658–4672, Oct. 2013.

[20] J. T. Parker, P. Schniter, and V. Cevher, ''Bilinear generalized approximate message passing—Part I: Derivation,'' *IEEE Trans. Signal Process.*, vol. 62, no. 22, pp. 5839–5853, Nov. 2014.

[21] A. P. Dempster, N. M. Laird, and D. B. Rubin, ''Maximum-likelihood from incomplete data via the *EM* algorithm,'' *J. Roy. Statist. Soc.*, vol. 39, no. 1, pp. 1–38, 1977.

[22] D. L. Donoho, A. Maleki, and A. Montanari, ''Message-passing algorithms for compressed sensing,'' *Proc. Nat. Acad. Sci. USA*, vol. 106, no. 45, pp. 18914–18919, Nov. 2009.

[23] H. Liu, X. Yuan, and Y. Zhang, ''Super-resolution blind channel-and-signal estimation for massive MIMO with one-dimensional antenna array,'' *IEEE Trans. Signal Process.*, to be published. doi: 10.1109/TSP.2019.2928977.

[24] J. A. Tropp and A. C. Gilbert, ''Signal recovery from random measurements via orthogonal matching pursuit,'' *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[25] D. Needell and J. A. Tropp, ''CoSaMP: Iterative signal recovery from incomplete and inaccurate samples,'' *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, 2009.

[26] L. Chen, A. Liu, and X. Yuan, ''Structured turbo compressed sensing for massive MIMO channel estimation using a Markov prior,'' *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4635–4639, May 2017.

[27] W. Yan and X. Yuan, ''Semi-blind signal detection for uplink massive MIMO with channel sparsity,'' in *Proc. IEEE Int. Conf. Commun.*, Shanghai, China, May 2019, pp. 1–12.

WENJING YAN received the B.S. degree in electronic and information engineering from Chongqing University, Chongqing, China, in 2018. She is currently pursuing the M.S. degree in electronic and communication engineering with the University of Electronic Science and Technology of China, Chengdu, China.

XIAOJUN YUAN (S'04–M'09–SM'15) received the Ph.D. degree in electrical engineering from the City University of Hong Kong, in 2008.

From 2009 to 2011, he was a Research Fellow with the Department of Electronic Engineering, City University of Hong Kong. He was also a Visiting Scholar with the Department of Electrical Engineering, University of Hawaii at Manoa, in 2009, as well as in the same period of 2010. From 2011 to 2014, he was a Research Assistant Professor with the Institute of Network Coding, The Chinese University of Hong Kong. From 2014 to 2017, he was an Assistant Professor with the School of Information Science and Technology, ShanghaiTech University. He is currently a Professor with the Center for Intelligent Networking and Communications, University of Electronic Science and Technology of China, supported by the Thousand Youth Talents Plan, in China. He has published over 130 peer-reviewed research papers in the leading international journals and conferences in the related areas. His research interests include broad range of signal processing, machine learning, and wireless communications, including but not limited to multi-antenna and cooperative communications, sparse and structured signal recovery, Bayesian approximate inference, and network coding. He has served on a number of technical programs for international conferences. He has been an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS, since 2017, and the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, since 2018. He was a Co-Recipient of the Best Paper Award of the IEEE International Conference on Communications (ICC) 2014, and a Co-Recipient of the Best Journal Paper Award of the IEEE Technical Committee on Green Communications and Computing (TCGCC) 2017.

• • •